





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 12.3758 seconds  
(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-12  
Perfect score: 1987  
Sequence: 1 ALLTSPYKVLARLIMWL.....RGVAKAVDFIPVESMETTR 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 283366 segs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	ID	Description
1	1902	95.7	3010	1 GNMVCI	genome polypotein
2	1890	95.1	3010	1 A45573	genome polypotein
3	1870	94.1	3010	1 GNMVTV	genome polypotein
4	1845	92.9	3010	1 S18030	genome polypotein
5	1823	91.7	3010	1 GNMVTC	genome polypotein
6	1715	86.3	3011	1 GNMVVC	genome polypotein
7	1709	86.0	3011	1 S40770	genome polypotein
8	1703	85.7	3011	1 GNMVCH	genome polypotein
9	1482	74.6	3014	1 UT5620	genome polypotein
10	1391	70.0	3033	1 GQ1303	genome polypotein
11	1391	70.0	3033	1 GNMVVC	genome polypotein
12	408.5	20.6	3005	2 T08841	polypotein - marm
13	342.5	17.2	2970	2 T08839	polypotein - marm
14	112	5.6	692	2 H71426	hypothetical prote
15	102.5	5.2	660	2 H71426	structural prote
16	101	5.1	564	2 S36637	signal recognition
17	101	5.1	600	2 B46642	DNA-directed DNA p
18	100.5	5.1	353	2 B46642	conserved hypotet
19	99	5.0	399	2 AH3038	conserved hypotet
20	99	5.0	399	2 C98247	hypothetical 50.8K
21	97.5	4.9	1085	2 T03531	coBN protein homol
22	95.5	4.8	2796	2 JG4098	tetracycline 6-lyd
23	94.5	4.8	2796	2 JG4743	fatty acid synthas
24	93.5	4.7	1380	2 AF1509	conserved hypotet
25	93.5	4.7	1380	2 T18309	receptor-adenylate
26	93	4.7	7463	2 T36248	CDA peptidase precu
27	92.5	4.7	716	2 G83612	transferrin precu
28	92.5	4.7	716	2 G83612	hypothetical prote
29	92.5	4.7	3069	2 H70656	fatty-acid synthas

30.	92	4.6	659	1 B44212	structural protein
31	91.5	4.6	428	1 AF0241	probable coenzyme
32	91.5	4.6	665	2 D83252	nucleotide sugar e
33	91.5	4.6	3414	1 GNMVNE	genome polypotein
34	91	4.6	434	2 G82728	conserved hypotet
35	91	4.6	904	2 A84212	hypothetical prote
36	90.5	4.6	868	2 H81775	aconitate hydratase
37	90.5	4.6	871	2 H72597	hypothetical prote
38	89.5	4.5	961	2 AE0375	probable cation-tr
39	88.5	4.5	401	1 A36961	pinin biogenesis p
40	88.5	4.5	446	2 AE1150	conserved hypotet
41	88.5	4.5	652	2 C70688	probable nitrate r
42	88.5	4.5	2121	2 T27406	hypothetical prote
43	88.5	4.5	3076	2 A87058	fatty acid synthas
44	87.5	4.4	3412	1 GNMVTV	genome polypotein
45	87	4.4	418	2 H90679	probable transport

ALIGNMENTS

RESULT 1  
GNMVCI genome polypotein - hepatitis C virus (strain J)  
N/contains: capsid protein C; envelope protein E; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
C/Accession: A39253; PS0086  
R/Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto  
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A/Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
A/Reference number: A39253; MUID:91088550; PMID:2175903  
A/Accession: A39253  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <KAT>  
A/Cross-references: GB:ID90208; NID:9221610; PIDD:BA14233.1; PIDD:9221611  
R/Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A/Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari  
A/Reference number: PS0085  
A/Accession: PS0086  
A/Molecule type: genomic RNA  
A/Residues: 2650-2707 <KAT>  
A/Experimental source: Japanese isolate  
A/Comment: The cleavage sites of this polypotein have not been determined.  
C/Species: hepatitis C virus genome polypotein  
C/Keywords: AFP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypotein; serin  
F/2115/Product: capsid protein C #status predicted <CP>  
F/116-191/Product: envelope protein M #status predicted <EPV>  
F/192-189/Product: major envelope protein E #status predicted <ME>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F/1230-1327/Region: nucleotide-binding motif A (P-loop)  
F/1112-1317/Region: nucleotide-binding motif B  
F/1116-1319/Region: DEKH motif  
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,325,417,423,430,448,532,556,623,645,1213,1255,2041,2077,2240,2

Query Match	95.7%	Score 1902;	DB 1;	Length 3010;
Best Local Similarity	94.2%	Pred. No. 6.9e-150;		
Matches 357;	Conservative 12;	Mismatches 10;	Indels 0;	Gaps 0;
OY	2	LLTSPYKVLARLIMWLQYITRVAHLQVWIPPLNVRGGDAIILLTCAVHPLIFD	61	
DB	828	LLTSPYKVLARLIMWLQYITRVAHLQVWIPPLNVRGGDAIILLTCAVHPLIFD	887	
OY	62	ITKLLALRPLNVLQAGIKVYFVRAGLIRACLVKRAAGHYVQNAFMKALATGT	121	
DB	888	ITKLLALRPLNVLQAGIKVYFVRAGLIRACLVKRAAGHYVQNAFMKALATGT	947	

QY 122 YVYHLLPPLQDMAAGLDLAVAVEPIFSDMEVXITTWGADTAACDIIISGLPVSARRG 181  
DB 948 YVYHLLPPLQDMAAGLDLAVAVEPIFSDMEVXITTWGADTAACDIIISGLPVSARRG 1007  
QY 182 REILLGADNFEQGGWMLLAPITAYSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241  
DB 1008 KEILLGRADSGEGGMLLAPITAYSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 1067  
QY 242 SFLATCNGVCWTVFHAGSKTLAGPKGPIITQYTNVDDLVGMQAPPGARSMTPTCTCGS 301  
DB 1068 SFLATCNGVCWTVFHAGSKTLAGPKGPIITQYTNVDDLVGMQAPPGARSMTPTCTCGS 1127  
QY 302 SDLYVTRHADVPVRRRGGSGSLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 361  
DB 1128 SDLYVTRHADVPVRRRGGSGSLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 1187  
QY 362 GYAKAVDPIPVESMETTR 380  
DB 1188 GYAKAVDPIPVESMETTR 1206

## RESULT 2

A45573  
genome polyprotein - hepatitis C virus (strain J1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: A45573  
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,  
Virus Res. 23, 39-53, 1992  
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  
A:Reference number: A45573; MUID:9225744; PMID:118627  
A:Accession: A45573  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3010 <TRAN>  
A:Cross-references: GB:D11168; GB:D01171; NID:9221612; PIDN:BA01943.1; PID:9221613  
A:Experimental source: HCV-JT  
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPW>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 95.1%; Score 1890; DB 1; Length 3010;  
Best Local Similarity 94.2%; Pred. No. 6,9e-149;  
Matches 357; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLIMWLOYLITRVAHLQWIPPLNVRGGDAIILLTCAVHPELFD 61  
DB 828 LTLSPYKYLARLIMWLOYLITRVAHLQWIPPLNVRGGDAIILLTCAVHPELFD 887  
QY 62 ITKLLAIFPLMWLOAGITKVYFVRAQGLIRACMLVKAAGHYVQMAFKALALGT 121  
DB 888 ITKLLAIFPLMWLOAGITKVYFVRAQGLIRACMLVKAAGHYVQMAFKALALGT 947  
QY 122 YVYHLLPPLQDMAAGLDLAVAVEPIFSDMEVXITTWGADTAACDIIISGLPVSARRG 181  
DB 948 YVYHLLPPLQDMAAGLDLAVAVEPIFSDMEVXITTWGADTAACDIIISGLPVSARRG 1007  
QY 182 REILLGADNFEQGGWMLLAPITAYSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241

DB 1008 REILLGADNFEQGGWMLLAPITAYSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 1067  
QY 242 SFLATCNGVCWTVFHAGSKTLAGPKGPIITQYTNVDDLVGMQAPPGARSMTPTCTCGS 301  
DB 1068 SFLATCNGVCWTVFHAGSKTLAGPKGPIITQYTNVDDLVGMQAPPGARSMTPTCTCGS 1127  
QY 302 SDLYVTRHADVPVRRRGGSGSLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 361  
DB 1128 SDLYVTRHADVPVRRRGGSGSLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 1187  
QY 362 GYAKAVDPIPVESMETTR 380  
DB 1188 GYAKAVDPIPVESMETTR 1206

## RESULT 3

GNVVTM  
genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40244  
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 188, 102-113, 1992  
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
A:Reference number: A40244; MUID:92230206; PMID:11314449  
A:Accession: A40244  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <CHE>  
A:Cross-references: GB:M64754  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructural  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPW>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 94.1%; Score 1870; DB 1; Length 3010;  
Best Local Similarity 92.1%; Pred. No. 3.2e-147;  
Matches 349; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLIMWLOYLITRVAHLQWIPPLNVRGGDAIILLTCAVHPELFD 61  
DB 828 LTLSPYKYLARLIMWLOYLITRVAHLQWIPPLNVRGGDAIILLTCAVHPELFD 887  
QY 62 ITKLLAIFPLMWLOAGITKVYFVRAQGLIRACMLVKAAGHYVQMAFKALALGT 121  
DB 888 ITKLLAIFPLMWLOAGITKVYFVRAQGLIRACMLVKAAGHYVQMAFKALALGT 947  
QY 122 YVYHLLPPLQDMAAGLDLAVAVEPIFSDMEVXITTWGADTAACDIIISGLPVSARRG 181  
DB 948 YVYHLLPPLQDMAAGLDLAVAVEPIFSDMEVXITTWGADTAACDIIISGLPVSARRG 1007  
QY 182 REILLGADNFEQGGWMLLAPITAYSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241  
DB 1008 REILLGADNFEQGGWMLLAPITAYSOOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 1067  
QY 242 SFLATCNGVCWTVFHAGSKTLAGPKGPIITQYTNVDDLVGMQAPPGARSMTPTCTCGS 301  
DB 1068 SFLATCNGVCWTVFHAGSKTLAGPKGPIITQYTNVDDLVGMQAPPGARSMTPTCTCGS 1127



QY 302 SDLYLVRHADVI PVRRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
DB 1128 SDLYLVRHADVI PVRRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 1187  
QY 362 GVAKAVDFIPVESMETTR 380  
DB 1188 GVAKAVDFIPVESMETTR 1206

## RESULT 4

SI8030  
genome polyprotein - hepatitis C virus (isolate JKI)  
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Variety: isolate JKI  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
C:Accession: SI8030; S33570; A48332; SI8029  
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991  
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie  
A:Reference number: SI8028  
A:Accession: SI8030  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <HON>  
A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479  
A:Experimental source: isolate JKI from an individual  
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
A:Reference number: A48332; MUID:93119270; PMID:8380322  
A:Accession: S33570  
A:Molecule type: genomic RNA  
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'D', 655-761, 'T', 763-782 <HON>  
A:Cross-references: EMBL:X61591  
A:Note: this sequence is inconsistent with the nucleotide translation  
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320  
as Trp, and TTC for residue 771 as Ser  
A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:P.121748)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepacivirin #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif B  
F/1316-1319/Region: DEXH motif  
F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 92.9%; Score 1845; DB 1; Length 3010;  
Best Local Similarity 91.8%; Pred. No. 3.9e-145;  
Matches 349; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 1 ALTLSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILITCAVHPELIF 60  
DB 827 ALTLSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILITCAVHPELIF 886  
QY 61 DITLLAIIFGLPMLVLOAGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 120  
DB 887 DITLLAIIFGLPMLVLOAGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 946  
QY 121 TYVVDHLTPLODMWAHAGLRDLAAVEVITFSDMEVKIITWGAIDPAAGDIIISGLPVARR 180  
DB 947 TYVVDHLTPLODMWAHAGLRDLAAVEVITFSDMEVKIITWGAIDPAAGDIIISGLPVARR 1006  
QY 181 GREILLGPADNFBGQWRLLAPITAYSQOTRGLGCIITSLTGRDKVQGEVQVSTAT 240

DB 1007 GREILLGPADNFBGQWRLLAPITAYSQOTRGLGCIITSLTGRDKVQGEVQVSTAT 1066  
QY 241 QSFLATCVNGVCMVTFHGAGSKTLAGPKGPIITQWYTNVDDVLVGMQAPPGARSMTPTCTG 300  
DB 1067 QSFLATCVNGVCMVTFHGAGSKTLAGPKGPIITQWYTNVDDVLVGMQAPPGARSMTPTCTG 1126  
QY 301 SSDLYLVRHADVI PVRRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCT 360  
DB 1127 SSDLYLVRHADVI PVRRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCT 1186  
QY 361 GVAKAVDFIPVESMETTR 380  
DB 1187 GVAKAVDFIPVESMETTR 1206

## RESULT 5

GNMVTIC  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
C:Accession: A38465  
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J  
U. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from human  
A:Reference number: A38465; MUID:91140698; PMID:1847440  
A:Accession: A38465  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <TKX>  
A:Cross-references: EMBL:MS8335; NID:G329770; PIDN:AA12945.1; PID:G329771  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: major envelope protein E #status predicted <MEB>  
F/192-389/Product: major envelope protein E #status predicted <NS1>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepacivirin #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif B  
F/1316-1319/Region: DEXH motif  
F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 91.7%; Score 1823; DB 1; Length 3010;  
Best Local Similarity 91.0%; Pred. No. 2.7e-143;  
Matches 345; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 2 LITLSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILITCAVHPELIF 61  
DB 828 LITLSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILITCAVHPELIF 887  
QY 62 ITKLLAIIFGLPMLVLOAGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 121  
DB 888 ITKLLAIIFGLPMLVLOAGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 947  
QY 122 YVVDHLTPLODMWAHAGLRDLAAVEVITFSDMEVKIITWGAIDPAAGDIIISGLPVARR 181  
DB 948 YVVDHLTPLODMWAHAGLRDLAAVEVITFSDMEVKIITWGAIDPAAGDIIISGLPVARR 1007  
QY 182 RETLLGPADNFBGQWRLLAPITAYSQOTRGLGCIITSLTGRDKVQGEVQVSTAT 241  
DB 1008 RETLLGPADNFBGQWRLLAPITAYSQOTRGLGCIITSLTGRDKVQGEVQVSTAT 1067  
QY 242 SFLLATCVNGVCMVTFHGAGSKTLAGPKGPIITQWYTNVDDVLVGMQAPPGARSMTPTCTG 301  
DB 1068 SFLLATCVNGVCMVTFHGAGSKTLAGPKGPIITQWYTNVDDVLVGMQAPPGARSMTPTCTG 1127  
QY 302 SDLYLVRHADVI PVRRGDSRGSLLSPRVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361

Db 1128 SDLVLRHADVIVRRRGDSRGLSPRPVSYLKSSGGPFLCPFGHVGIFRAVCTR 1187  
 QY 362 GVAADVPIPVESMETTMR 380  
 Db 1188 GVAADVPIPVESMETTMR 1206

## RESULT 6

GNWVC3

genome polypeptide - hepatitis C virus (strain HCV-1)  
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C/Accession: A39166; PMID:91172826; PMID:1848704

R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A/Title: Genetic organization and diversity of the hepatitis C virus.  
 A/Reference number: A39166; PMID:91172826; PMID:1848704

A/Accession: A39166  
 A/Molecule type: mRNA  
 A/Residues: 1-3011 <CHO>

A/Cross-references: GB:M62221; NID:G229873; PIDN:AAA5676.1; PID:G229874

R/Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to

A/Reference number: P00393; PMID:92268871; PMID:1316939  
 A/Accession: P00403

A/Molecule type: genomic RNA

A/Residues: 1577-1633 <CHA>

A/Cross-references: DDBJ:D10128

A/Experimental source: isolates E-b16

A/Accession: P00404

A/Status: preliminary

A/Molecule type: genomic RNA

A/Residues: 1577-1633 <CH2>

A/Experimental source: isolates E-b17

C/Superfamily: hepatitis C virus genome polypeptide

C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F/1-115/Product: capsid protein C #status predicted <CPC>

F/116-191/Product: major envelope protein M #status predicted <EMP>

F/192-389/Product: major envelope protein E #status predicted <MEP>

F/390-1206/Product: nonstructural protein NS1 #status predicted <NS1>

F/1207-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F/1230-1237/Region: nucleotide-binding motif A (P-loop)

F/1312-1317/Region: nucleotide-binding motif B

F/1316-1319/Region: DEXH motif

F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>

F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>

F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 86.3%; Score 1715; DB 1; Length 3011;  
 Best Local Similarity 82.8%; Pred. No. 2.8e-134;  
 Matches 313; Conservative 32; Mismatches 33; Indels 0; Gaps 0;

QY 3 LTLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFDI 62  
 Db 829 LTLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFDI 888  
 QY 63 TKLLAIFGPMVLQGLITKVPYFVRAOGLIRACMLVRGAAGHYVQMAFMKLAALTGTY 122  
 Db 889 TKLLAIFGPMVLQGLITKVPYFVRAOGLIRACMLVRGAAGHYVQMAFMKLAALTGTY 948  
 QY 123 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWAGDTAACGDIISGLPVSARRGR 182  
 Db 949 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWAGDTAACGDIISGLPVSARRGR 1008  
 QY 183 EILIGPADNFEQGMRLAPITAYSOQTRGLGCIITSLTGDKXQVGEVQVSTATOS 242  
 Db 1009 EILIGPADNFEQGMRLAPITAYSOQTRGLGCIITSLTGDKXQVGEVQVSTATOS 1068

QY 243 FLATCNGVCTVFFHAGSKTLAGPKPITOMYTNVDDLVGMQAPPGAREMTPTCTGSS 302  
 Db 1069 FLATCNGVCTVFFHAGSKTLAGPKPITOMYTNVDDLVGMQAPPGAREMTPTCTGSS 1128  
 QY 303 DLVLTFRADVIPIVRRRDSRGLSPRPVSYLKSSGGPFLCPFGHVGIFRAVCTR 362  
 Db 1129 DLVLTFRADVIPIVRRRDSRGLSPRPVSYLKSSGGPFLCPFGHVGIFRAVCTR 1188

QY 363 GVAADVPIPVESMETTMR 380  
 Db 1189 GVAADVPIPVESMETTMR 1206

## RESULT 7

S40770

genome polypeptide - hepatitis C virus

N/Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 19-Jan-2001  
 C/Accession: S40770; PMID:1210721

R/Okamoto, H.  
 submitted to the EMBL Data Library, March 1992

A/Reference number: S40770

A/Accession: S40770

A/Molecule type: genomic RNA

A/Residues: 1-3011 <OKA>

A/Cross-references: EMBL:D10749; NID:G221586; PIDN:BA00705.1; PID:G221587

R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Un. J. Exp. Med. 60, 167-177, 1990

A/Title: The 5'-terminal sequence of the hepatitis C virus genome.

A/Reference number: F1284; PMID:9101316; PMID:210721

A/Accession: F1285

A/Molecule type: genomic RNA

A/Residues: 1-513 <OK2>

A/Cross-references: GB:D00831; NID:G221511; PIDN:BA00705.1; PID:G221512

A/Experimental source: isolate HC-J1

C/Superfamily: hepatitis C virus genome polypeptide

C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine

F/2-115/Product: capsid protein C #status predicted <CPC>

F/116-191/Product: major envelope protein M #status predicted <EMP>

F/192-389/Product: major envelope protein E #status predicted <MEP>

F/390-1206/Product: nonstructural protein NS1 #status predicted <NS1>

F/1207-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F/1230-1237/Region: nucleotide-binding motif A (P-loop)

F/1312-1317/Region: nucleotide-binding motif B

F/1316-1319/Region: DEXH motif

F/1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>

F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>

F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 86.0%; Score 1709; DB 1; Length 3011;  
 Best Local Similarity 82.3%; Pred. No. 8.8e-134;  
 Matches 311; Conservative 32; Mismatches 35; Indels 0; Gaps 0;

QY 3 LTLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFDI 62  
 Db 829 LTLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFDI 888  
 QY 63 TKLLAIFGPMVLQGLITKVPYFVRAOGLIRACMLVRGAAGHYVQMAFMKLAALTGTY 122  
 Db 889 TKLLAIFGPMVLQGLITKVPYFVRAOGLIRACMLVRGAAGHYVQMAFMKLAALTGTY 948  
 QY 123 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWAGDTAACGDIISGLPVSARRGR 182  
 Db 949 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWAGDTAACGDIISGLPVSARRGR 1008  
 QY 183 EILIGPADNFEQGMRLAPITAYSOQTRGLGCIITSLTGDKXQVGEVQVSTATOS 242  
 Db 1009 EILIGPADNFEQGMRLAPITAYSOQTRGLGCIITSLTGDKXQVGEVQVSTATOS 1068  
 QY 243 FLATCNGVCTVFFHAGSKTLAGPKPITOMYTNVDDLVGMQAPPGAREMTPTCTGSS 302

Db 1069 FLATCINGCWTYHAGAGRTTASPKGPIQWTNTVDOLVGMPPAGRSRLTPTCTGSS 1128  
QY 303 DLYVTRHADVPYRRGDSRGSLLSPRPSYLLKSSGGGELLCPGSHAVGIFRAAVCTRG 362  
Db 1129 DLYVTRHADVPYRRGDSRGSLLSPRPSYLLKSSGGGELLCPGSHAVGIFRAAVCTRG 1188  
QY 363 VAKAVDFIPVESMETTMR 380  
Db 1189 VAKAVDFIPVESMETTMR 1206

## RESULT 8

GNMVCH  
genome polypeptide - hepatitis C virus (strain H)  
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
A/Note: host Homo sapiens (man)  
C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C/Accession: A36814; A41546  
R:Inchouse, G.; Zebeder, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Submitted to Genbank, July 1992  
A/Description: Genomic structure of the human prototype strain H of hepatitis C virus:  
A/Reference number: A36814  
A/Accession: A36814  
A/Molecule type: genomic RNA  
A/Residues: 1-3011 <INC>  
A/Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
R:Inchouse, G.; Zebeder, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
A/Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
A/Reference number: A41546; MUID:92052256; PMID:1658800  
A/Contents: annotation  
A/Note: neither amino acid nor nucleotide sequence is given  
C/Superfamily: hepatitis C virus genome polypeptide  
C/Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructural  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EMP>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitisin #status predicted <NS3>  
F:1230-1237/Product: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEKH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 85.7%; Score 1703; DB 1; Length 3011;  
Best Local Similarity 82.3%; Pred. No. 2.8e-13;  
Matches 311; Conservative 34; Mismatches 33; Indels 0; Gaps 0;  
QY 3 LTLSPYVYVLLARLIMWLQYLTTRVEALQWVIPPPLNVRGGRDAIILLTCVHPELIPDI 62  
Db 829 LTLSPYVYVLLARLIMWLQYLTTRVEALQWVIPPPLNVRGGRDAIILLTCVHPELIPDI 888  
QY 63 TKLLAATGPIPLMTLQAGITKPYFVRAQGLIRACLVKKAAGHYVQMAFMKLAALTSGT 122  
Db 889 TKLLAATGPIPLMTLQAGITKPYFVRAQGLIRACLVKKAAGHYVQMAFMKLAALTSGT 948  
QY 123 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRGR 182  
Db 949 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRGR 1008  
QY 183 ELLIGPADNFEQGGRLAPITAYAOQTRGLIGCITISLTGRDKQVGEVGVSTANOS 242  
Db 1009 ELLIGPADNFEQGGRLAPITAYAOQTRGLIGCITISLTGRDKQVGEVGVSTANOS 1068  
QY 243 FLATCNGCWTYHAGAGRTTASPKGPIQWTNTVDOLVGMPPAGRSRLTPTCTGSS 302

Db 1069 FLATCINGCWTYHAGAGRTTASPKGPIQWTNTVDOLVGMPPAGRSRLTPTCTGSS 1128  
QY 303 DLYVTRHADVPYRRGDSRGSLLSPRPSYLLKSSGGGELLCPGSHAVGIFRAAVCTRG 362  
Db 1129 DLYVTRHADVPYRRGDSRGSLLSPRPSYLLKSSGGGELLCPGSHAVGIFRAAVCTRG 1188  
QY 363 VAKAVDFIPVESMETTMR 380  
Db 1189 VAKAVDFIPVESMETTMR 1206

## RESULT 9

JCS620  
genome polypeptide - hepatitis C virus (isolate EUH1480)  
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C/Accession: JCS620  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, J.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A/Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
A/Reference number: JCS620; MUID:97366593; PMID:9223423  
A/Accession: JCS620  
A/Molecule type: mRNA  
A/Residues: 1-3014 <CHN>  
A/Cross-references: GB:Y13184  
A/Experimental source: genotype 5a, which predominates in South Africa  
A/Note: the translation of the nucleotide sequence is not complete in this paper  
C/Superfamily: hepatitis C virus genome polypeptide  
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EMP>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:384-408/Region: hypervariable #status predicted  
F:590-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: hepatitisin #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEKH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interfeon sensitivity determining #status predicted

Query Match 74.6%; Score 1482; DB 1; Length 3014;  
Best Local Similarity 69.4%; Pred. No. 7.7e-115;  
Matches 263; Conservative 56; Mismatches 60; Indels 0; Gaps 0;  
QY 2 LTLSPYVYVLLARLIMWLQYLTTRVEALQWVIPPPLNVRGGRDAIILLTCVHPELIPDI 61  
Db 829 LTLSPYVYVLLARLIMWLQYLTTRVEALQWVIPPPLNVRGGRDAIILLTCVHPELIPDI 888  
QY 62 TKLLAATGPIPLMTLQAGITKPYFVRAQGLIRACLVKKAAGHYVQMAFMKLAALTSGT 121  
Db 889 TKLLAATGPIPLMTLQAGITKPYFVRAQGLIRACLVKKAAGHYVQMAFMKLAALTSGT 948  
QY 122 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRGR 181  
Db 949 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRGR 1008  
QY 182 ELLIGPADNFEQGGRLAPITAYAOQTRGLIGCITISLTGRDKQVGEVGVSTANOS 241  
Db 1009 ELLIGPADNFEQGGRLAPITAYAOQTRGLIGCITISLTGRDKQVGEVGVSTANOS 1068  
QY 242 FLATCNGCWTYHAGAGRTTASPKGPIQWTNTVDOLVGMPPAGRSRLTPTCTGSS 301  
Db 1069 FLATCNGCWTYHAGAGRTTASPKGPIQWTNTVDOLVGMPPAGRSRLTPTCTGSS 1128  
QY 302 SDLYVTRHADVPYRRGDSRGSLLSPRPSYLLKSSGGGELLCPGSHAVGIFRAAVCTRG 361  
Db 1129 SDLYVTRHADVPYRRGDSRGSLLSPRPSYLLKSSGGGELLCPGSHAVGIFRAAVCTRG 1188

QY 362 GVAKAVDFIPVESMETMR 380  
 DB 1189 GVAKALEFVPEVNETTMR 1207

## RESULT 10

J01303  
 genome polyprotein - hepatitis C virus (isolate HC-J6)  
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome having poor homology to repository NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Kura, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A/Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
 A/Reference number: J01303; MUID:92044440; PMID:1658196  
 A/Accession: J01303  
 A/Molecule type: genomic RNA  
 A/Residues: 1-303 <OKA>  
 A/Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651  
 A/Experimental source: isolate HC-J6 from a Japanese individual  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; trans  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: major envelope protein M #status predicted <EMP>  
 F/192-389/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/390-733/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/734-1010/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1011-1619/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1620-1866/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,233,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 70.0%; Score 139.1; DB 1; Length 3033;  
 Best Local Similarity 66.0%; Pred. No. 3e-107;  
 Matches 250; Conservative 55; Mismatches 74; Indels 0; Gaps 0;

QY 2 LTLSPYKYLALRLWLYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELLD 61  
 DB 832 LFTLTPGYKTLRLRFWMWLCYLLTLAEAWQEWAPPMQVRGGRDGIIMAVAFCEGVVD 891  
 QY 62 ITKLTAIFGRLVQAGITKVPYFRAQGLIRACMLYKKAAGHYVQAKPKLAALTG 121  
 DB 892 ITKMLAVIGRAVLKAGLALTRVYFRAHLLRMCTMVAHLAGRYVQAVLALGRWGT 951  
 QY 122 YVVDHITPLQDWAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRG 181  
 DB 952 YVVDHITPLQDWAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRG 1011  
 QY 182 RETLLPAPNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 241  
 DB 1012 RETLLPAPNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 1071  
 QY 242 SPLATCVMGCVTFVHAGSKTLAPKGPITQMTNNVDQDLYGWAAPGASMTCTCGS 301  
 DB 1072 SPLATCVMGCVTFVHAGSKTLAPKGPITQMTNNVDQDLYGWAAPGASMTCTCGS 1131  
 QY 302 SDLYLTVRADVTPVRRRDSRGLSPRPVSYLKGSSGGLPLCSGNAVGIFRAAVCT 361  
 DB 1132 SDLYLTVRADVTPVRRRDSRGLSPRPVSYLKGSSGGLPLCSGNAVGIFRAAVCT 1191  
 QY 362 GVAKAVDFIPVESMETMR 380  
 DB 1192 GVAKSIDFIPVETLIVR 1210

RESULT 11  
 GNNVJB

genome polyprotein - hepatitis C virus (strain HC-J8)  
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome having poor homology to repository NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 R/Okamoto, H.; Kura, K.; Okada, S.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;  
 Virol. 188, 331-341, 1992  
 A/Title: Full-length sequence of a hepatitis C virus genome having poor homology to repository  
 A/Reference number: A40250; MUID:92230232; PMID:1314459  
 A/Accession: A40250  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3033 <OKA>  
 A/Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609  
 R/Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing  
 A/Reference number: P00397; MUID:92268871; PMID:1316939  
 A/Accession: P00397  
 A/Molecule type: genomic RNA  
 A/Residues: 2678-2754 <CHN>  
 A/Cross-references: DDBJ:D10134  
 A/Experimental source: isolate B-b12  
 R/Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,  
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
 A/Title: Distribution of plural HCV types in Japan.  
 A/Reference number: P00554; MUID:92068204; PMID:1720309  
 A/Accession: P00554  
 A/Molecule type: mRNA  
 A/Residues: 2678-2729 <KAT>  
 A/Cross-references: GB:D10562; GB:D90518; NID:G221523; PIDN:BA01418.1; PID:G221524  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F/1-115/Product: capsid protein C #status predicted <CPC>  
 F/116-191/Product: major envelope protein M #status predicted <EMP>  
 F/192-389/Product: nonstructural protein NS1 #status predicted <NS1>  
 F/390-733/Product: nonstructural protein NS2 #status predicted <NS2>  
 F/734-1010/Product: nonstructural protein NS3 #status predicted <NS3>  
 F/1011-1619/Product: nonstructural protein NS4 #status predicted <NS4>  
 F/1620-1866/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F/1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
 F/196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23;

Query Match 70.0%; Score 139.1; DB 1; Length 3033;  
 Best Local Similarity 64.2%; Pred. No. 3e-107;  
 Matches 244; Conservative 64; Mismatches 72; Indels 0; Gaps 0;

QY 1 ALTLSPYKYLALRLWLYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELLD 60  
 DB 831 SIFLTPAYKILSLRSYTWLSTYVLAEOIQWVPLEVRGGRDGIIMAVAFCEGVVD 890  
 QY 61 DITLLAIFGRLVQAGITKVPYFRAQGLIRACMLYKKAAGHYVQAKPKLAALTG 120  
 DB 891 EITKMLAIGRAVLKAGLALTRVYFRAHLLRMCTMVAHLAGRYVQAVLALGRWGT 950  
 QY 122 YVVDHITPLQDWAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRG 180  
 DB 951 YVVDHITPLQDWAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASRRG 1010  
 QY 182 RETLLPAPNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 240  
 DB 1011 RETLLPAPNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 1070  
 QY 242 SPLATCVMGCVTFVHAGSKTLAPKGPITQMTNNVDQDLYGWAAPGASMTCTCGS 300  
 DB 1071 SPLATCVMGCVTFVHAGSKTLAPKGPITQMTNNVDQDLYGWAAPGASMTCTCGS 1130  
 QY 302 SDLYLTVRADVTPVRRRDSRGLSPRPVSYLKGSSGGLPLCSGNAVGIFRAAVCT 360  
 DB 1132 SDLYLTVRADVTPVRRRDSRGLSPRPVSYLKGSSGGLPLCSGNAVGIFRAAVCT 1190

Db 1131 AVDLIVITMNAVDPYRKDDRGALLSPRLSTLKGSSGCVLGRHAGVLFRAAVCA 1190  
 QY 361 RGVAAKADFIPIVSEMETTNR 380  
 Db 1191 RGVAAKSIDFIPIVSELDVATR 1210

RESULT 12

T08841  
 polyprotein - douroucouli hepatitis GB virus A  
 C/Species: douroucouli hepatitis GB virus A  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C/Accession: T08841  
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08841  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 13005 <ERK>  
 A/Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40501.1; PID:g2828600  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 20.6%; Score 408.5; DB 2; Length 3005;  
 Best Local Similarity 31.3%; Pred. No. 2.9e-25;  
 Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 41 RGRD-----AIIILCAVHPELIFDITKLLAIFGLMLOAGITKYPYVRAQGLRAC 96  
 Db 813 RGRDWMVTVVWVAAGIIFREVRCGALTA-----LAALISIDYLETL-ILTA 864  
 QY 97 MLVRKA-----AGHYVQMAFMKLAALGTYYVDHLPLQDMAAGRLDAVAVEPY 148  
 Db 865 QPAAALDLSTFLGADITRAFRLEKRGVTLFQHCQVXGAAIIXDGLALEY 924  
 QY 149 IFSMEVKITTWGADTAACGDIISGLFVSARREIILG--PADNEGGQWMLLAPITAY 206  
 Db 925 SVIARDCVIRADARTLACGQRVEGLVVARRGDEVLVGFPSRALPFGVPTAPVYV- 983  
 QY 207 SQCTRGILGCIITSLTRDXNQVGEVQVNSTAQSLACVNGVCMVFHGGSKTLAG 266  
 Db 984 MORLGFPSVYKTSMLRDRHEHGSIVLGTSTRSMGTCVGVMTTFHGSNAKTLAG 1043  
 QY 267 PKGPIQWYTVNDLVLGWMQAPGARSMPTCTGSSDLVYVTHADYIPVRRGDSRGSJ 326  
 Db 1044 PVGVNCRWSPSDVAIVPLPSASCLPECKCTGQVWCIRN--DGALCHGRLSKLVEL 1101  
 QY 327 LSPRPVSYLKGSGGGLCPGSGHAAGIFRAAVCTRGV-----AKAVDPIVES 374  
 Db 1102 DLPTFISDFRSGSSPILCDEGHVGMV-VSVLHRGVKVTGVRYVKEWETLPKDS 1155

RESULT 13

T08839  
 polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: T08839  
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <ERK>  
 A/Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 17.2%; Score 342.5; DB 2; Length 2970;

Best Local Similarity 28.9%; Pred. No. 9.1e-20;  
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;

QY 65 LLALIFG-----LMVLQAGITKYPYFRAQGLLRACMLYKAKAGHYVQMAFMYKA 116  
 Db 816 LVAAFMFMRRIAAVCATILGFGFFVDVYLEVVLSSPVLRLAVLSLVAAGDKLA 875  
 QY 117 ALGTYYVDHLPLQD--MAAG-----LRDLAVAVEPIFSMEVKITTWGADTA 165  
 Db 876 T---TWLVEKLRKKNCFLYAAGVTRTRTAQLQWGFALPEVAVHEDCAMVDAARTL 932  
 QY 166 ACGDIISGLPSARREIILGPDADNPEGQWRL-----LAPITANSQCTRGILGCIIT 219  
 Db 933 SCQGSVHKKPVVARSGEVLGLVNGV---WELPFGFVTPAPVAVH-HHKGSGFGVYKT 987  
 QY 220 SLTGRDXNQVGEVQVNSTAQSLFATCVNGVCMVFHGGSKTLAGFKPITQWYTVND 279  
 Db 988 SMTGMDERHGNVYVLTSTRSMGTCVGVMTTFHGSNAKTLAQMGPVNSRWMGAS 1047  
 QY 280 QDLVGMQAPPGARSMTPTCTGSSDLVYVTHADYIPVRRGDSRGSLS----- 328  
 Db 1048 DDAVAYPLPVGAKCLEPCKCOPQGVWVI-----RND--GALCHGTIGRTVELDL 1094  
 QY 329 PRPVSYLKGSGGGLCPGSGHAAGIFRAAVCTRG-----VAKAVDPIVSEMETT 378  
 Db 1095 PAELCDFRSGSSPILCDEGHVGMV-LSVLHRSRVTGIRYTKPMWETLPREAITHT 1150

RESULT 14

H71426  
 hypothetical protein - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 A/Variety: Columbia  
 C/Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
 R/Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirke  
 P.; Medler, H.; Medler, S.; Kotter, P.; Ertian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
 Nature 391, 485-488, 1998  
 A/Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,  
 erioft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichman, S.; Ansc  
 C.; Chalmers, N.  
 A/Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
 A/Reference number: A71400; MUID:98121113; PMID:9461215  
 A/Accession: H71426  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-692 <BEV>  
 A/Cross-references: GB:Z97340; NID:g2244950; PID:e327492; PID:g2244965  
 C/Genetics:  
 A/Map position: 4COP9-4G3845

Query Match 5.6%; Score 112; DB 2; Length 697;  
 Best Local Similarity 24.1%; Pred. No. 0.25;  
 Matches 97; Conservative 45; Mismatches 135; Indels 126; Gaps 23;

QY 13 LARLIWQLYITRVAHQVWIPPLNVGSGDAIILTCVHPELIFDITKLLAIFGP 72  
 Db 83 IADLAGIYVLR-----QGNFPRASVYAGSNCIEL-----KGPILMDLTFLR-FLT 131  
 QY 73 LMYLQAGITKVPY--FVRAQGLIRACMLYR--KAAAGHYVQMAFM----- 113  
 Db 132 LCMV---FSKKPFAVLEBAGYTHEDVLLQKPKAGGHTMOPAFITIRDTNSKICILLIR 188  
 QY 114 -----KLAALGTYYVDHLPLQDMAAGRLDAVAVEPIFSMEVKITTWGADTA 166  
 Db 189 GHISIDCTIATAGVAPPHSHVLRH--GGLSNVLVLR-----AH 226  
 QY 167 CGDIISGLPSARREIILGPDADNPEGQW--RLAPITAVSQCTRGILGCIITSLTGR 224  
 Db 227 CG-----WVAAR-----WIAKLSVP-----DLTKAL-- 248  
 QY 225 DXNQVGEVQVNSTA---TQSFATCV-----NGVCMVTFHGGSKTLAGFKGP-I 271

```

Db      249 DENV-SERVOIVGSHLGGGTASLLTYILRBOKEFASATCTFAAGTPTNLMINGESGKHF 307
QY      272 TQMTYNDODLV--GWQAPPGASMTPTCGSSDLVYTRHADVIP-VRRGDSRGLL 327
Db      308 TTIING--SDLVPTFSASSVDLRSSEVTSSWSNDLRDQVHTRVLSVTKSATATGSR 365
QY      328 SPRPVSYLKSSGGPILCP--SGHAVGIFRAAVCTRGVAKAVD 368
Db      366 PSINAKAKAVAGAILRPVSSGTQVAFLVNGC--GKIKCID 406

```

## RESULT 15

```

VHMMH2
structural protein 2 precursor - hepatitis E virus (strain Burma)
C/Species: hepatitis E virus
C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999
C/Accession: C40778
R/Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.
Virology 185, 120-131, 1991
A/Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi
A/Reference number: A40778; MUID:92024067; PMID:1926770
A/Accession: C40778
A/Molecule type: genomic RNA
A/Residues: 1-660 <TRM>
A/Cross-references: GB:W73218; NID:9330023; PIND:AAA5736.1; PID:9330026
C/Note: the authors translated the codon Cgc for residue 2 as Ala
C/Superfamily: hepatitis E virus structural protein 2
C/Keywords: structural protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-660/Product: structural protein 2 #status predicted <SP2>

```

Query Match 5.2%; Score 102.5; DB 1; Length 660;

Best Local Similarity 19.8%; Pred. No. 1.5; Mismatches 130; Indels 151; Gaps 19;

```

Db      63 TKLL--AIFGPMVLQAG-----ITKPYEVR--AQLIRACMLYKAAAGHYVQVA 111
QY      151 TNLVLAAPLSPLEPQDGTNTIMATASNYAQVRAATIRYRLVPMNAVGYAISTIS 210
Db      112 FKLALITGYVDHLTPLODMAHAGRLDAVAPEVIPSDEVKITWGADTAACGDI 171
QY      211 FWPQTTTPTSV-----DNNSITSTVRILVQPIASELVI----- 246
Db      172 SGLPVASARGREILLGPAD--NFEQGGWRLAPI--TAYSQOTRGL-----GCITSLTG 223
QY      247 -----PSERLHTRNQGMRSVETSGVAEEBATSGLVMLCTHGLVNSYTN 290
Db      224 -----RDKNQVEGEVQVNSTATQSF 244
QY      291 TPTGALGLDFALELEFNNLTPGNTNTRVSRYSTARRHLRGADGTALTTTATRFM 350
Db      245 A----TCVAGV-----CMTVZH-----GAG-----SKTLAHPKG-PIT 272
QY      351 KDLFTSTNGVGEIGIALTLTNLADTLGLPTELLISSAGQLFYSRPVVSANGEPV 410
Db      273 QMTYNDODLVGMAPEPGASMTPTCGSSDLVY--TRHADVIPVRRGDSRG-SLIS 328
QY      411 KLYTSVENA---QQDKGIALPHDIDIGBSRVVLDYDNQHEQDRPTSPASRRPFSYLR 466
Db      329 PRPVSYLK-----GSSGGPILCSGHAVGIFRAAVCTRGVAKAVDFIPY 372
QY      467 ANDVLMISLTAAYDOSTYSSSTGPVYV--SDSVTLVNVATGAQAVARSLDMTKV 519

```

Search completed: May 6, 2004, 09:37:18  
Job time : 13.3758 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.9319 Seconds

(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-12

Sequence: 1 ALLTSPYKVLARIWL.....RGVAKAVFIPVSMETMR 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1502	95.7	3010	1	POLG_HCVUA
2	1890	95.1	3010	1	POLG_HCVUT
3	1870	94.1	3010	1	POLG_HCVTW
4	1823	91.7	3010	1	POLG_HCVBK
5	1715	86.3	3011	1	POLG_HCVI
6	1703	85.7	3011	1	POLG_HCVH
7	1391	70.0	3033	1	POLG_HCVU8
8	1391	70.0	3033	1	POLG_HCVU6
9	102.5	5.2	660	1	VST2_HEVHU
10	102.5	5.2	660	1	VST2_HEVVA
11	101	5.1	564	1	SR5C_ARAVH
12	101	5.1	600	1	DP02_MOUSE
13	93.5	4.7	1380	1	CYAA_LEIPO
14	93	4.7	704	1	TOLB_CHLRE
15	92.5	4.7	434	1	TREB_HORSE
16	92	4.6	659	1	VST2_HEVME
17	91.5	4.6	485	1	VST2_HEVHH
18	91.5	4.6	660	1	VST2_HEVHY
19	91.5	4.6	3414	1	POLG_TBEVW
20	91	4.6	444	1	V447_XYLFA
21	90.5	4.6	3414	1	POLG_TBEVH
22	89.5	4.5	961	1	ATCU_YERRE
23	89	4.5	444	1	V327_XYRPT
24	88.5	4.5	401	1	PILC_PSEBU
25	87.5	4.4	3412	1	POLG_TBSVS
26	86.5	4.4	347	1	MDHM_EUCBU
27	86.5	4.4	470	1	MDHM_IAMVM
28	86	4.3	338	1	GALE_NEIGO
29	86	4.3	433	1	DCUA_WOLSU
30	86	4.3	441	1	MDHP_MESCR
31	85	4.3	470	1	NRAM_TAPRA
32	85	4.3	730	1	HEIS_MEYNA
33	85	4.3	834	1	PWP2_SCHPO

34	85	4.3	3313	1	CLR3_RAT	088278	rattus norv
35	84.5	4.3	1705	1	PTPV_MOUSE	P70289	mus musculus
36	84	4.2	309	1	UCP2_RAT	P56500	rattus norv
37	84	4.2	339	1	GPDA_COREF	Q8FDP0	corynebacte
38	84	4.2	403	1	MHP1_ECOLI	P75889	escherichia
39	84	4.2	423	1	MRO9_MOUSE	Q9WU6	mus musculus
40	84	4.2	470	1	NRAM_IARBE	P31510	influenza a
41	84	4.2	503	1	V402_MYCTU	005586	mycobacteri
42	83.5	4.2	538	1	DAC_FCTSP	P39045	actinomadu
43	83.5	4.2	1399	1	RPOC_PSEAB	Q3W69	pseudomonas
44	83	4.2	341	1	MDHM_BRANA	Q43744	brassica na
45	82.5	4.2	453	1	NHAC_BACSU	007553	bacillus su

## ALIGNMENTS

RESULT 1  
POLG\_HCVUA STANDARD; PRT; 3010 AA.  
ID POLG\_HCVUA  
AC P26662;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein (contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadnavirina) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).  
DE Hepatitis C virus (isolate Japanese) (HCV).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
OC NCBI\_TaxID=11116;  
RX MEDLINE=9108550; PubMed=2175903;  
RA Kato N., Hijioka M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;  
RT "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
RP [2]  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijioka M., Nakagawa M., Ootsuyama Y., Muraishi K., Ohkoshi S., Shimotohno K.;  
RT "Molecular structure of the Japanese hepatitis C viral genome.";  
RL FEBS Lett. 280:325-328(1991).  
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N').  
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and RNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use, by non-profit institutions as long as its content is in no way modified and that statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL, D90208; BA014233.1; -

DR PIR; A99253; GNNVCJ.  
 DR HSSD; P26663; IUXP.  
 DR MEROPS; S29.001; -.  
 DR MEROPS; U39.001; -.  
 DR InterPro; IPR0009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002516; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; Peptidase C29.  
 DR InterPro; IPR007095; RNA\_pol\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSV1r.  
 DR Pfam; PF01545; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF0186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC1.  
 DR Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 95.7%; Score 1902; DB 1; Length 3010;

Best Local Similarity 94.2%; Pred. No. 6,9e-148;  
 Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 2 ILTSPYKVLARLIWVQYLITREAHLOWIPPLNVGGRDAIILTCVAPELIFD 61  
 DB 828 ILTSPYKVFARLIWVQYLITREAHLOWIPPLNVGGRDAIILTCVAPELIFD 687  
 QY 62 ITKLLALFGLPMVQAGITVPRVRAQGLIRACMLVRAAGHYVOMAKLAALTGT 121  
 DB 888 ITKLLALFGLPMVQAGITVPRVRAQGLIRACMLVRAAGHYVOMAKLAALTGT 947  
 QY 122 YVYDHLTFLRMAHAGLADLVAVEPVFSDMEVKIITWAGDTAACGDIISGLPVSARG 181  
 DB 948 YVYDHLTFLRMAHAGLADLVAVEPVFSDMEVKIITWAGDTAACGDIISGLPVSARG 1007  
 QY 182 REILIGPDNREGGCKMLAITYASQTRGLGCIITSLTRKKNVGEVQVSTATQ 241  
 DB 1008 KEILIGPDNREGGCKMLAITYASQTRGLGCIITSLTRKKNVGEVQVSTATQ 1067  
 QY 242 SFELATVAGVCMVTHGAGSKTLGPKPIQMTVNDODLVGQAPGARSMPTCTGS 301  
 DB 1068 SFELATVAGVCMVTHGAGSKTLGPKPIQMTVNDODLVGQAPGARSMPTCTGS 1127  
 QY 302 SDLYLVTRHADVIPRRRGDSGLSPRPVSYLKSGSGGLLCPSGHAYGIFRAVCTR 361  
 DB 1128 SDLYLVTRHADVIPRRRGDSGLSPRPVSYLKSGSGGLLCPSGHAYGIFRAVCTR 1187  
 QY 362 GVAAKADPIPVESMETTR 380  
 DB 1188 GVAAKADPIPVESMETTR 1206

RESULT 2  
 ID POLG\_HCVUT STANDARD; PRT; 3010 AA.  
 AC 000259;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis virus)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-JT) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CC NCBI\_Taxid=31642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92295714; PubMed=1318627;  
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.,  
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
 RT carrier: sequence variation within the same individual and among  
 RT infected individuals.";  
 RL Virus Res. 23:39-53(1992).  
 RL -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Glu in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipid-protein envelope. The envelope consists of two proteins: a  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.





CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M84754; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A40244; GNMVTW.  
 DR PDB; 1NS3; 25-FEB-03.  
 DR PDB; 1NS3; 08-APR-98.  
 DR MEROPS; S29.001; -;  
 DR MEROPS; U39.001; -;  
 DR InterPro; IPR009003; Cys\_Ser\_tryptase.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NSI.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSV1r.  
 DR Pfam; PF01543; HCV\_core; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KM Polypeptide; Transferase; RNA-directed RNA polymerase;  
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KM 3D-structure.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 209  
 REMOVED FROM CAPSID PROTEIN C BY THE  
 CELLULAR AMINOPEPTIDASE.  
 CORE PROTEIN (POTENTIAL).  
 MATRIX PROTEIN (POTENTIAL).  
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 PROTEASE/HELICASE NS3 (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 NONSTRUCTURAL RNA POLYMERASE (POTENTIAL).  
 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 POTENTIAL.  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 ATP (POTENTIAL).  
 DECH\_BOX.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;  
 Query Match 94.1%; Score 1870; DB 1; Length 3010;  
 Best Local Similarity 92.1%; Pred. No. 3e-145;  
 Matches 349; Conservative 16; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 LITLSPYKVLALILMWLYLITRVAHLQWVIPPINVRGGRDAIILTCVAPEDILFD 61  
 DB LITLSPHKMFLALILMWLYLITRVAHLQWVIPPINVRGGRDAIILTCVAPEDILFD 887  
 QY 62 ITKLALIFGLPLVLOAGITVPPFVRAQGLIRACMLVRAAGHYVQMAFMKLAALGT 121  
 DB ITKLALIFGLPLVLOAGITRIPFVRAQGLIRACMLVRAAGHYVQMAFMKLAALGT 947  
 QY 122 YVYDHLFPLQMAHAGRDIAVAEPIFSDMEKILITWGAADTAACGDIISGLPVSARRG 181  
 DB YVYDHLFPLQMAHAGRDIAVAEPIFSDMEKILITWGAADTAACGDIISGLPVSARRG 1007  
 QY 948 YVYDHLFPLQMAHAGRDIAVAEPIFSDMEKILITWGAADTAACGDIISGLPVSARRG 1007  
 DB YVYDHLFPLQMAHAGRDIAVAEPIFSDMEKILITWGAADTAACGDIISGLPVSARRG 1007  
 QY 182 REILGPADNEEGGKMLLAPITAYSQOTRGLLGCITSLTGRDKNVESEVOYVSTATQ 241  
 DB REILGPADNEEGGKMLLAPITAYSQOTRGLLGCITSLTGRDKNVESEVOYVSTATQ 1067  
 QY 242 SFLATCNGVCWTFPHAGSKTLAGPKPIITOWTANTYDOLVGMQAPRGRSMTPCTCGS 301  
 DB SFLATCNGVCWTFPHAGSKTLAGPKPIITOWTANTYDOLVGMQAPRGRSMTPCTCGS 1127  
 QY 302 SDLYLVTRHADVIVRRRCDSRGLSPRPVSYLKGSSGGPILCPSSHAAGIFRAAVCTR 361  
 DB SDLYLVTRHADVIVRRRCDSRGLSPRPVSYLKGSSGGPILCPSSHAAGIFRAAVCTR 1187  
 QY 1128 SDLYLVTRHADVIVRRRCDSRGLSPRPVSYLKGSSGGPILCPSSHAAGIFRAAVCTR 1187  
 DB SDLYLVTRHADVIVRRRCDSRGLSPRPVSYLKGSSGGPILCPSSHAAGIFRAAVCTR 1187  
 QY 362 GVAKAVDFIPVSEMETMR 380  
 DB GVAKAVDFIPVSEMETMR 1206  
 RESULT 4  
 POLG\_HCVK STANDARD; PRT; 3010 AA.  
 AC P26663;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin).  
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate BK) (HCV).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_Taxid=11105;  
 RN [1]





```

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NF_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FC5A99 CRC64;

Query Match 86.3%; Score 1715; DB 1; Length 3011;
Best Local Similarity 82.8%; Pred. No. 1,6e-132;
Matches 313; Conservative 32; Mismatches 33; Indels 0; Gaps 0;

3 LTLSPYKVLARLAWLQYLITREVAHQVMIPLNVRGSDATLLTCAVHELIPTDI 62
829 LTLSPYKRYISWCLMWLQYFLTRVEAQHLVMIPLNVRGSDAVILLCAVHELIPTDI 888
63 TKLLIAIFGLMVLQGITVPEFVAQGLIRACMVRXAAGHYVOMAFMKLAALTGY 122
889 TKLLIAIFGLMVLQGITVPEFVAQGLIRACMVRXAAGHYVOMAFMKLAALTGY 948
123 VYDHLPLQDMANAGADLAVAEPIVFSDEMKIITWGAADTAACDIIISGLPVSARRR 182
949 VYHMLPLRLMAMHNRDLAVAEPIVFSDEMKIITWGAADTAACDIIISGLPVSARRR 1008
183 EILLGPADNEGGQMLARITAYSOOTRGLICITSLGRPKNOVEGVQVVSATQS 242
1009 EILLGPADNEGGQMLARITAYSOOTRGLICITSLGRPKNOVEGVQVVSATQS 1068
243 FLATCVGVQWMTVEFGAGSKTLGPKPITOMTYNDQDLVGMQAPPGASMPCTCGSS 302
1069 FLATCVGVQWMTVEFGAGSKTLGPKPITOMTYNDQDLVGMQAPPGASMPCTCGSS 1128
303 DLYLVTRHADVTPYRRRGGSRGSLSPRPVSYIKGSGGGLLPSCGHAVIFPAAYCTG 362
1129 DLYLVTRHADVTPYRRRGGSRGSLSPRPVSYIKGSGGGLLPSCGHAVIFPAAYCTG 1188
363 VARAVDFIPVESMETWR 380
1189 VARAVDFIPVESMETWR 1206

```

```

RESULT 6
POLG HCVA STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);

```

```

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepatitis C
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate H) (HCV).
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
DE Hepacivirus.
OC NCBI_TaxID=11108;
RN (1)
RN SEQUENCE FROM N.A.
RN MEDLINE=92052256; Pubmed=1658800;
RN Inchauspe G., Zebadee S., Lee D.H.H., Sugtani M., Nasoff M.,
RN Prince A.M.;
RN "Genomic structure of the human prototype strain H of hepatitis C
RN virus: comparison with American and Japanese isolates."
RN Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).
RN (2)
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RN MEDLINE=97331322; Pubmed=9187654;
RN Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RN "Structure of the hepatitis C virus RNA helicase domain."
RN Nat. Struct. Biol. 4:463-467 (1997).
RN (3)
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RN MEDLINE=98154321; Pubmed=9493270;
RN Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RN Murcko M.A., Lin C., Caron P.R.;
RN "Hepatitis C virus NS3 RNA helicase domain with a bound
RN oligonucleotide: the crystal structure provides insights into the mode
RN of unwinding."
RN Structure 6:89-100 (1998).
RN (1) FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
RN (2) FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
RN NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
RN (3) FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
RN ACTIVATION OF NS3.
RN (4) FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVITY ROLE.
RN (5) FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
RN ESSENTIAL ROLE IN THE VIRUS REPLICATION.
RN (6) CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
RN precursor polypeptide; commonly with Asp or Glu in the P6
RN position, Cys or Thr in P1 and Ser or Ala in P1'.
RN (7) CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
RN {RNA} (N).
RN (8) SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
RN LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
RN AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
RN (9) PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
RN PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
RN (10) SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
RN (11) SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
RN (12) This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN (13) EMBL; M67463; AAA45534.1; -.
RN PIR; A36814; GNVVCH.
RN PDB; 1HRI; 25-NOV-98.
RN PDB; 1AIV; 16-FEB-99.
RN PDB; 1AIR; 17-JUN-98.
RN MEROPS; S29.001; -.
RN MEROPS; U39.001; -.
RN TRNSPAC; T04155; -.
RN InterPro; IPR009003; Cys_ser_trypsin.
RN InterPro; IPR001410; DEAD.
RN InterPro; IPR002522; HCV_capsid.

```



DB 1129 DLTAVTRHADVIRERRGRSGSLSPRISYKSGSGPLLCPTGHAVALGRAVACTRG 1188  
 QY 363 VAKAVDFIPRESMETMR 380  
 DB 1189 VAKAVDFIPVENLETMR 1206

RESULT 7  
 POLG\_HCVU6 STANDARD; PRT: 3033 AA.  
 ID POLG\_HCVU6  
 AC P26660;  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP33); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J6) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2004440; Pubmed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kuzai K., Lizuka H.,  
 RA Machida A., Miyakawa Y., Mayumi M.,  
 RT Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704(1991).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and RNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL, D00944; BAA00792.1; -  
 CC PIR, J01303; J01303.  
 CC HSSP, P27958; 1HE1.  
 CC MEROPS, S29.001; -  
 CC InterPro: IPR009003; Cys\_Ser\_tryptsin.  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV\_capsid.  
 CC InterPro: IPR002521; HCV\_core.  
 CC InterPro: IPR002519; HCV\_env.  
 CC InterPro: IPR002531; HCV\_NS1.  
 CC InterPro: IPR002518; HCV\_NS2.  
 CC InterPro: IPR000745; HCV\_NS4A.  
 CC InterPro: IPR001490; HCV\_NS4B.  
 CC InterPro: IPR002868; HCV\_NS5A.  
 CC InterPro: IPR002166; HCV\_RdRP.  
 CC

DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4A; 1.  
 DR Pfam: PF01001; HCV\_NS4B; 1.  
 DR Pfam: PF01506; HCV\_NS5A; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PolyProtein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1330 1333  
 FT CARBOHYD 136 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 224 224  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 FT SEQUENCE 3033 AA; 329165 MW; F957F5CA273B93E CRC64;  
 SO

Query Match 70.0%; Score 1391; DB 1; Length 3033;  
 Best Local Similarity 66.0%; Pred. No. 7, 1e-106;  
 Matches 250; Conservative 55; Mismatches 74; Indels 0; Gaps 0;

QY 2 LTLSPYVYVLLARLILWLTQYITLREVALQWIRPLNVRGGRDAIILITCAVHELFID 61  
 DB 832 LFTLTPGYKTLISRFILWMLCYLITLAEAVQWAPMQVGRDGLIWAVALFCGEVFD 891  
 QY 62 ITGLIAIFGPIMLVLOAGITKYVEFVRAQGLIRACGLYKKAAGHYVQWAFKLAALTGT 121  
 DB 892 ITKMLAIVGPRAYLKGALTRIPYFIRAHALLRMCTMVRHLAAGRVQVLLALGRWTGT 951  
 QY 122 YVYDHLTPQDVAHAGRLDAVAVEPIVSDMEVKIITGADTAACGDIISGLPVSARAG 181  
 DB 952 YIYDHLTPMSDWAANGLRDLAVALVEPIIPSPMEKKVIWGAETAAAGDILHGLPVSARLG 1011

QY 162 REILGSPADNPEQGGWLLAPITAYSOOTRGLICITLSTGRDKNQVEGVWSTATQ 241  
 DB 1012 REVLGADGDTSGKMLLPITRAYAOITGLLTIVSNKTKGDKIQAEIIVLSTVQ 1071  
 QY 242 SFATCNVGVCMVTFHAGSXTLAGPKPITOMTVNDOLVMOAPPGARSMPTCTGCS 301  
 DB 1072 SFLGTTISGLVMTVYHAGNKTLAGSGPVTOMVSSAEGDLVGMPSPPGKSLPECTCGA 1131  
 QY 302 SDLYLVTRADVYFVRKRGSRGSLSPRVSVTLKSGSGPFLCPSGHAIVGIRAVCTR 361  
 DB 1132 VDLVTVRNADVIFPARRRGKRGALLSPRLSTLKSGSGPVLCPRGHAGVFAAVCSR 1191  
 QY 362 GVAKAVDFIPVESMETMR 380  
 DB 1192 GVAKSIDFIPVETLIDIVTR 1210

RESULT 8  
 POLG\_HCVU8 STANDARD; PRT; 3033 AA.  
 AC P2665;  
 ID POLG\_HCVU8  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein  
 DE NS5B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCBI\_TaxID=11115;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9230232; PubMed=1314459;  
 RA Okamoto H., Kuzai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes.";  
 RL Virology 188:331-341(1992).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC 110protein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D10988; BAA01761.1; -  
 DR PIR; A40250; GMVU8.  
 DR HSP; P27958; IHEI.  
 DR MEROPS; S29.001; -  
 DR MEROPS; U39.001; -  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4A.  
 DR InterPro; IPR001490; HCV\_NS4B.  
 DR InterPro; IPR002868; HCV\_NS5A.  
 DR InterPro; IPR002166; HCV\_NS5B.  
 DR InterPro; IPR004109; peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVLR.  
 DR Pfam; PF01542; HCV capsid; 1.  
 DR Pfam; PF01543; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4A; 1.  
 DR Pfam; PF01001; HCV\_NS4B; 1.  
 DR Pfam; PF01506; HCV\_NS5A; 1.  
 DR Pfam; PF00998; viral\_RBP; 1.  
 DR Pfam; PF018662; HCV\_NS1; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural  
 DR INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT SITE 1087 1087  
 FT ACT SITE 1111 1111  
 FT ACT SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7B381FD1A CRC64;

Query Match 70.0%; Score 1391; DB 1; Length 3033;  
 Best Local Similarity 64.2%; Pred. No. 7; le-106;  
 Matches 244; Conservative 64; Mismatches 72; Indels 0; Caps 0;  
 QY 1 ALTLSPYKYLARLIMVQYLTRVEALQWVPPVLRGGRDAIILTCVHRLIF 60  
 DB 831 SIFLTAYKILLRSYVWLIVLAQIQWVPLVEVGGRDGIWVAVILFRLVF 890



[illegible]

```

RESULT 9
VST2_HEVBU
ID_VST2_HEVBU STANDARD; PRT; 660 AA

```

DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Structural protein 2 precursor (ORF2).  
OS Hepatitis E virus (strain Burma) (HEV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage;  
OC Hepatitis E-like viruses.  
NCBI\_TaxID=31767;  
[1]  
RD SEQUENCE FROM N.A.  
RX MEDLINE=92024067; PubMed=1926770;  
RA Tam A.W., Smith M.W., Guerra M.E., Huang C.-C., Bradley D.W.,  
FA Fry K.E., Reyes G.R.;  
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the  
RT full-length viral genome.";  
RL Virology 185:120-131(1991).  
-!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR	EMBL	Accession	DR	EMBL	Accession
DR	PIR	C40778; VHWMA2.	DR	PIR	C40778; VHWMA2.
DR	InterPro	IPR004261; SP2.	DR	InterPro	IPR004261; SP2.
DR	InterPro	IPR008975; Viral_cap_coat.	DR	InterPro	IPR008975; Viral_cap_coat.
DR	Pfam	PF03014; SP2; 1.	DR	Pfam	PF03014; SP2; 1.
KW	Signal.		KW	Signal.	
FT	SIGNAL	1..19	FT	SIGNAL	1..19
FT	CHAIN	20..660	FT	CHAIN	20..660
FT	STRUCTURAL	POTENTIAL	FT	STRUCTURAL	POTENTIAL
FT	PROTEIN	PROTEIN 2.	FT	PROTEIN	PROTEIN 2.
SEQUENCE	660 AA; 70978 MW; 58352A013CC04A61C CR664;		SEQUENCE	660 AA; 70978 MW; 58352A013CC04A61C CR664;	

Query Match	5.2%	Score 102.5;	DB 1;	Length 660;
Best Local Similarity	19.8%;	Pred. No. 0.93;		
Matches 82; Conservative	52;	Mismatches 130;	Indels 151;	Gaps 19

QY 63 TKLL--AIFGPLMLQAG-----ITKVPYVR--AQGLIRACMLVRKAAGHYVQMA 111

```

Db      151  TNLVLNAPLSPILLPLQDGTNTHTIMATEASNYAQRYVARATIRPRPLVPNAVGGYAISTS 210
Qy      112  FMKIALITGYVYDHLTPLODMHAGRLDLAVNEPIIFSDMEVKIITWGDATACGDI 171
Db      211  FWPOTITTFTSV-----DMNSTIDVAILVQFGLASLV-----246
Qy      172  SGLVARSARGREILLGPAD--NFEQGWRLLAPI--TAVSQGTRELL-----GCIIITSLTG 223
Db      247  -----PSERLHYANQGMRSVETSVAEEBEXTSGVLMLCIRGSLVNSYTN 290
Qy      224  -----RDKNQVEGVQVVSATATQSEL 244
Db      291  TPTYGALGLDFALEFELRPNTNTVRVSSTARHRRRGADGIAELITTPATRFM 350
Qy      245  A-----TCVNGV-----CWTVFH-----GAG-----SKTLAGEKG--PIT 272
Db      351  KDLYFTSNGVGETGRGAILTFRLADTLGLPTLISAGGGLFVSRPVSANGBEPTV 410
Qy      273  QMTYNVDDLVKQWQAPPGARSMPCTGSSPLLIV--TRADAVIPARRBEDSRG--SLLS 328
Db      411  KLYTSVENA-----QODKGIALPHDIDGESRVYVQDYDNQEDQRPFPSPARSFRFSYLR 466
Qy      329  PRPYSYXK-----GSSGGLPCPSHAYGIFPAAYCTRGVAKAVDFIPV 372
Db      467  ANDYVIMLSLTAEBVDOSTYSSSTGPVYV--SDSTVLVNVAVGAGAVARASLDWTKV 519

```

RESULT 10	
VST2_HEVPA	
ID VST2_HEVPA	STANDARD;
	PRT: .660 AA

DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Structural protein 2 precursor (ORF2).  
OS Hepatitis E virus (strain Pakistan) (HEV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage;  
OC Hepatitis E-like viruses.  
OX NCBI\_TaxID=33774;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92115700; PubMed=1731327;  
RA Tsareva S.U., Emerson S.U., Reyes G.R., Tsareva T.S., Legeers L.J.,  
RA Walk I.A., Iqbal M., Purcell R.H.;  
RT "Characterization of a prototype strain of hepatitis E virus."  
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992)  
CC - FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sb.ch/announce>  
CC or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).

DR	EMBL; M80581; AAA45727.1; -.	
DR	InterPro; IPR004261; SP2.	
DR	InterPro; IPR008975; Viral_cap_coat.	
DR	Pfam; PF03014; SP2; 1.	
KW	Signal.	
FT	SIGNAL.	1 22 BY SIMILARITY.
FT	CHAIN	23 660 STRUCTURAL PROTEIN 2.
SQ	SEQUENCE	660 AA; 70980 MW; 8085BC53C9B46FD3 CRC64;

Query Match	5.2%	Score 102.5	DB 1	Length 660
Best local similarity	19.8%	Pred. No. 0.93		
Matches 62	Conservative	54	Mismatches 128	Indels 151
				Gaps 20

63 TKLL--AIFGLMVLQAG-----ITKVPYFVR--AQGLIRACMLVRKAAGGHVQMA 111

```

DQ 151 TNLVLAAPSPPLPDDGNTNTHIMATASNAQYKARATITRPLVPAVAGVATISIS 210
QY 112 FPKLAALTYGVYDHLTPLODMAGRLDVAVEPVISDMKVIITWAGDPAACDII 171
DQ 211 FVQQTITPTSV-----DMNSITSDVIRLVQGIASELVI----- 246
QY 172 SGPLVARRRREILIGPAD--NFEQGMRL-----APINAYQ 208
DQ 247 -----PSERLHFRNQGMRSVETSVAEBEATISGLVMLCHGSPVASYN 290
QY 209 QT-RGLGCT-----ITSLTGRDNQ-----VEGEVQVSTATQSP 244
DQ 291 TPTGALGLDPALELEFRNLTPGNTNTRVRSYSTARRHLRGRAGDTALTTATATRM 350
QY 245 A-----TCVNGV-----CMTVH-----GAG-----SKTLAPKG-PIT 272
DQ 351 KDLYFTSTNGVGEIGRIALLTLNLDLTLGLPTELISAGQLFYSSPVSANSEPV 410
QY 273 QMTYNDQDILVQWQAPGARSMPTCGSSDLYLV---TRHADVIFVRSGSRG-SLIS 328
DQ 411 KLTYSVENA-----QDQKGIAPHIDLGESRVVIGQYDQHEQDRTPEPAPSRPSVLR 466
QY 329 PRVYSYK-----GSSGGLPCPSGHAAGIFRAAVCTRGVAKAVDFIV 372
DQ 467 ANDVLMSTPAEYDOSTYGSSTGPPYV--SDSVTLVNVATGAQAVARSLDWTXV 519

RESULT 11
SRSC_ARATH
ID SRSC_ARATH STANDARD; PRT; 564 AA.
AC P37107; O82570;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Signal recognition particle 54 kDa protein, chloroplast, precursor
DE (SRP54) (54 kDa chloroplast protein) (54CP) (PFC).
DE PFC OR AT5G03940 OR P8F6_150.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Streptophyta; eudicotyledons; core eudicots; rosids;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=94012817; PubMed=8408079;
RA Franklin A.E., Hoffman N.E.;
RT "Characterization of a chloroplast homologue of the 54-kDa subunit of
RT the signal recognition particle.";
RL J. Biol. Chem. 268:22173-22180(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;
RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,
RT encoding the 54 kDa subunit of chloroplast signal recognition
RT particle.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=1130714;
RA Tabara S., Karako T., Nakamura Y., Kotani H., Kato T., Asanuma E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naito K., Okumura S., Shino S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoenking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dancie M.,
RA Du H., Edwards J., Fryman J., Haakenen B., Lamer E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozerik P., Riley A., Strommatt C.,

```

```

RA Wagner-McPherson C., Molan A., Yeakum M., Bell M., Dedina N.,
RA Parnell L., Shah R., Rodriguez K., Hoon See L., Vli D., Baker J.,
RA Kiroff K., Toth K., King L., Bahret A., Miller B., Matra M.A.,
RA Marienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Mambert R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Berr E., Johnson S.,
RA Langham S.-A., McCullach B., Robben J., Grymompier B., Zimmermann W.,
RA Rampeger U., Medler H., Balke K., Medler E., Peters S.,
RA van Steeven T., Dirke W., Nothman P., Klein Lankhorst R.,
RA Weltzenegger T., Bothe G., Rose M., Haut J., Bernieris S., Hempel S.,
RA Feldpasch M., Lamberth S., Villarroel R., Giesen J., Adiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K., Rud S., Schoof H.,
RA Scheller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.,
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana".
RT Nature 408:823-826(2000).
CC -1- FUNCTION: May target chloroplast proteins to either the thylakoid
CC or envelope membranes.
CC -1- SUBCELLULAR LOCATION: Chloroplast stroma.
CC -1- TISSUE SPECIFICITY: Most abundant in green shoot tissue and
CC lower levels seen in the roots and etiolated buds.
CC -1- SIMILARITY: Belongs to the GTP-binding SRP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; Z21970; CAA79891.1; -
CC DR EMBL; AF092168; AAC64139.1; -
CC DR EMBL; AL162873; CAB85514.1; -
CC DR PIR; S36637; S36637.
CC DR HSSP; O07347; 1PFH.
CC DR InterPro; IPR003593; AAA_ATPase.
CC DR InterPro; IPR000897; SRP54.
CC DR InterPro; IPR004125; SRP54_SRP.
CC DR InterPro; IPR004780; SRP_sub.
CC DR Pfam; PF00448; SRP54_1.
CC DR Pfam; PF02881; SRP54_N_1.
CC DR Pfam; PF02978; SRP_SFBL_1.
CC DR SMART; SM00382; AAA_1.
CC DR TIGRFAMs; TIGR00959; ffh_1.
CC DR PROSITE; PS00300; SRP54_1.
CC KW Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;
CC TRANSIT peptide.
CC FT TRANSIT 1 75
CC FT CHAIN 76 564
CC FT DOMAIN 76 370
CC FT DOMAIN 371 564
CC FT NP_BIND 183 190
CC FT NP_BIND 265 269
CC FT NP_BIND 323 326
CC FT CONFIDCT 76 76
CC SQ SEQUENCE 564 AA; 61232 MW; 423F7285FB9063E4 CRC64;
Query Match 5.1%; Score 101; DB 1; Length 564;
Best Local Similarity 26.1%; Pred. No. 1;
Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;
QY 54 VHPRLI---FDITKLALIFGLPLVYLAQI-----TKYPTVRAQGLIRACMLVR 100
DQ 154 VHEELVLMGGEVSEILOPAKSGPTVILLGLQGVGKTTVCARAKLACYIKKQG--KSCMLI- 210
QY 101 KAAAGHVQAFAFKALAL---TGTYYVDHLTPLO--DMAHAGRLDVAVEPVISDMKVI 155
DQ 211 --ADVTTPRAIDVLYLGQVGVVPTVYAGTDVKAIDAKGKAKK-----NNVDV 261
QY 156 KIITWAGDTAACDIIISGSPVARRRREIL-----LGPADNFEQGMRLIAPITAYSQOT 210

```

Db 262 VIN-...STRGRUQIDKMKMDELKDYKFFLNPTVELLVDAATGQ--EAAALVTFNVEI 315  
 QY 211 RGLAGCITSLNGRDKNQVEGVQVVS 237  
 Db 316 -GITGAILTLTDGDSRGGAALSVKEVS 341

RESULT 12  
 DP02 MOUSE STANDARD; PRT; 600 AA.  
 ID DP02 MOUSE  
 AC P33611;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).  
 GN POLA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 84-102; 269-285 AND 394-403.  
 RX MEDLINE=93216788; PubMed=8463324;  
 RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Uti M.,  
 RA Hanaka F.;  
 RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA  
 RT polymerase alpha-primase complex and their gene expression during  
 RT cell proliferation and the cell cycle.";  
 RL J. Biol. Chem. 268:8111-8122(1993).  
 CC -1- FUNCTION: May play an essential role at the early stage of  
 CC chromosomal DNA replication by coupling the polymerase  
 CC alpha/primase complex to the cellular replication machinery (By  
 CC similarity).  
 CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme  
 CC (subunits A, B, C and D), which is assembled throughout the cell  
 CC cycle. The largest subunit (subunit A) has DNA polymerase  
 CC activity, the two smallest subunits (subunits C and D) have DNA  
 CC primase activity. Subunit B binds to subunit A.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M  
 CC PHASE (By SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D13546; BAA02746.1; -  
 DR PIR: B46642; B46642.  
 DR MGI: 996690; PolA2.  
 DR InterPro: IPR007200; DNA\_pol\_alpha\_B.  
 DR Pfam: PF04058; DNA\_pol\_alpha\_B; 1.  
 KW DNA replication; Nuclear protein; Phosphorylation.  
 FT DOMAIN 101 107 POLY-GU.  
 FT DOMAIN 115 157 PRO/SER/THR-RICH (HYDROPHILIC).  
 FT SEQUENCE 600 AA; 66267 MW; 79F9ABE6EF3FEEBC CRC64;

Query Match 5.1%; Score 101; DB 1; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 1.1;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 92 LIRACMIVKXAGHYOM-AFMKLAULT-----GTYYVDH-----TPQDMA 134  
 Db 27 LAELCVLRQEDGWSELNFCISAGKTCITVILNSFEYVINKLSAKMSASQDSG 86  
 QY 135 HAGRLDAVAVPVIFSDMEVKIITWGADTAACDI--ISGLP-----VSAHREGREI 184  
 Db 87 HAGTRDI-VSIQELIEAEHEEETLLSSVTTPSKGLPKRVASSTPTPLTKRSVAARSRC- 144

QY 185 ILGPADNREGGWRRLAITYASQOTRGLGCITSLTGRDNQVEGVQVVSATQSF 244  
 Db 145 LLSPPS-----FSPSAPPSQK-----YTSRTNR-----GEVVTTFGSAC-- 178  
 QY 245 ATCVAGVCMVTFHAGAGSKTL--AGPKGPITQMYTNVDDPLVG 284  
 Db 179 -----GLSMGRKSGSGSVSLKVVGPDEPLTGSYKAMFQGLMG 215

RESULT 13  
 ID CYAA LEIDO STANDARD; PRT; 1380 AA.  
 AC Q27675;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Receptor-type adenylylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-  
 DE lyase) (Adenylyl cyclase).  
 GN RAC-A.  
 OS Leishmania donovani.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=IS Sudanese;  
 RX MEDLINE=95340554; PubMed=7615561;  
 RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.,  
 RA "A family of putative receptor-adenylylate cyclases from Leishmania  
 RT donovani.";  
 RL J. Biol. Chem. 270:17551-17558(1995).  
 CC -1- FUNCTION: Could act as a receptor for a unknown ligand.  
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
 CC -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote)  
 CC but not in the mammalian host stage of the parasite life cycle.  
 CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U17042; AAA74998.1; -  
 DR PIR: T18309; T18309.  
 DR InterPro: IPR001054; G\_Cyclase.  
 DR Pfam: PF00211; guanylate\_cyc; 1.  
 DR SMART: SM00044; CYCC; 1.  
 DR PROSITE: PS50125; GUANYLATE CYCLASES\_2; 1.  
 KW Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein;  
 KW Metal-binding; Magnesium.  
 FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT DOMAIN 56 891 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 892 912 POTENTIAL.  
 FT DOMAIN 913 1380 CYTOPLASMIC (POTENTIAL).  
 FT METAL 938 938 MAGNESIUM (BY SIMILARITY).  
 FT METAL 981 981 MAGNESIUM (BY SIMILARITY).  
 FT CARBOHYD 422 432 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 497 497 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT SEQUENCE 1380 AA; 151692 MW; 6B2D5F7D3C1107A0 CRC64;

Query Match 4.7%; Score 93.5; DB 1; Length 1380;  
 Best Local Similarity 21.0%; Pred. No. 13;  
 Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;

QY 42 GGRDAIILLTCAVHP---ELIFDTIKL-----LIALFGPLM-----V 75

```

Db 101 GGRPIKIL-----HPPDQDNLYDIAEVILHSLAROEKLAIVGPIYDLGRLTAALSNADV 155
Qy 76 LQGITVP-----YFRAQGLIRACLVKAKAGHYV-----QMAFK 114
Db 156 VQSGMLIAFPSSGVRTSDSYFTRAEPMELKVLVLM-----HIVRLRRKRVAFNR 210
Qy 115 LAAATG-----TYVYDHLTPLODWAHAGLRDLA-VAPVIFSDMEVK-----156
Db 211 ---LTGHHFGSEELTYQDTLTSL-----LRDPALYLVFVSSESVDEHAFDAMAD 260
Qy 157 -----ITWGNADPAACDIIISGLFVBARREILG-----PADNFEQ 195
Db 261 TNPQVITVMAAPVQVYVFEKVLTPRISAVVICSMTQKRVDPVYKELLSAGSIKQ 320
Qy 196 GWELLAPITAYSOQTRGLLCITITSLTRDKNOVEG-----VQVSTATQSFATCVNGV 251
Db 321 DGRILASAT-----SPVSGELKYMVELKAMNSNIE---NSG 356
Qy 252 CMTVFHAGSKTLAGEP-----GPITQMTVNVDDLVGMQAPPARSMTCCTGSSDLVYV 307
Db 357 SFYYPDDSTETLGRARSEAPLSRKYT-VDE---FQAHPSIAKLMLGMLSTLVQ 412
Qy 308 TRHADVPVRRGDSKSLSPRPV---SYLKSSGSPILIC-PSGHAVGIFRAVCTRGV 363
Db 413 TLEQTKIVRSTYKGLFNGQNFVIGGCVLDYGGP---CEPLAGPLG--ASCYCNQGG 468
Qy 364 AKAV-----DFIPVESME 376
Db 469 HSSILTVLQNASWDIVDSSFK 490

RESULT 14
TOLE_CHLITE STANDARD; PRT; 434 AA.
ID TOLE_CHLITE
AC Q8KED0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE TOLE protein precursor.
GN TOLE OR CT0636.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobiaceae.
OX NCBI_TaxId=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUS / ATCC 49652 / DSM 12025;
RX MEDLINE=2103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Niernan W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Redune D.,
RA Vamathevan J., Knorr H., White O., Gruber T.W., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TJS, a
RT photosynthetic, anaerobic, green-sulfur bacterium."
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
CC -!- FUNCTION: Involved in the tonB-independent uptake of proteins (by
CC similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to the tolB family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC EMBL; AE012837; AMW71875.1;

```

```

DR TIGR; CT0636; -.
DR HAMAP; MF_00671; -, 1.
KT Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL. 1 27 Potential.
FT CHAIN. 28 434 TolB protein.
SQ SEQUENCE 434 AA; 47275 MW; FIA347B89C7A0F99 CRC64;

Query Match
Best Local Similarity 21.2%; Pred. No. 3.4;
Matches 69; Conservative 47; Mismatches 19; Indels 70; Gaps 14;

Qy 70 FGLPMTVQAGITKPYFVRAQGLIRACLVKAKAGHYVGMQAPFKALALNGTYVYDHLTP 129
Db 8 FALCITMFGMLFVFTLRABE-VGEYTAIRK-EGASRIAVLDKISADGKQ-----58
Qy 130 LQDMAH-----GLRDLA-VAPVIFSDMEVKIITWGNADPAACDIIISGLFVS 177
Db 59 -REWARSLDVTITNGKLPFTGLFNLPA--PLNHNQNGGLNFASIASVGGDIYAGGSVT 115
Qy 178 ARRGREIL-----LPPADNFEQGRRLAPITAYSOQTRGLLCITTS 220
Db 116 KRSGRPVLEMHVYDSSGKSLARTYTESQGLAIGLRFQADLVELLTKRSVFGTRIVF 175
Qy 221 LTGRDKNQ-----VEGEVQVSTATQSFATCV-----NGVCVTFHAGSKTLAPKG 269
Db 176 VANRTGKEIKMCDPDEENVQLNRSISLTPAVSPDGYIAMTDV--TSGKNLYIKN 233
Qy 270 PITQMTVNVDDLV-----GMQAPPARSMTCCTC--GSSDLVYTRHADVPVRRGDSR 323
Db 234 IATARKVSNMGHCISPAWR--PGITVTLVTLSEGDODLYLI--RADGTVERRLTKGG 289
Qy 324 GSLLSP-----RPVSYLKSSGSP 342
Db 290 GIDVSPITFSPDGSXMAVSTRGSGP 314

RESULT 15
TRFE_HORSE STANDARD; PRT; 706 AA.
ID TRFE_HORSE
AC P27425;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sero transferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal
DE binding globulin).
GN TF.
OS Equus caballus (Horse).
OC Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxId=9796;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277958; PubMed=8504171;
RA Carpenter M.A., Broad T.B.;
RT "The cDNA sequence of horse transferrin."
RL Biochim. Biophys. Acta 1173:230-232 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Extracellular tissue;
RA McQuell K.J., Adams M.H., Baker C.B.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transferrins are iron binding transport proteins which
CC can bind two atoms of ferric iron in association with the binding
CC of an anion, usually bicarbonate. It is responsible for the
CC transport of iron from sites of absorption and heme degradation to
CC those of storage and utilization. Serum transferrin may also have
CC a further role in stimulating cell proliferation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -!- DOMAIN: Composed of two homologous domains.
CC -!- SIMILARITY: Belongs to the transferrin family.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC The European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; U65020; AAA63058.1; -.  
DR EMBL; U21127; AAA63884.1; -.  
DR PIR; S33761; S33761.  
DR HSP; P02787; IABE.  
DR InterPro; IPR001156; Transferrin.  
DR Pfam; PFO0405; transferrin\_2.  
DR PRINTS; PRC0422; TRANSFERRN.  
DR SMART; SMO0094; TRFER; 2.  
DR PROSITE; PS00205; TRANSFERRIN 1; 2.  
DR PROSITE; PS00206; TRANSFERRIN 2; 2.  
DR PROSITE; PS00207; TRANSFERRIN 3; 2.  
DR Transport; Iron transport; Glycoprotein; Metal-binding; Repeat.  
KW Signal.

```
Gy      110 MAENKLAALGTGYV---YDHLPLODMAHAGRIDLAVAVEPYFSDSNEVKIITWGA----    162
          ::::|::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      321 LGEFRLRPAMDWTLYLGVEYT-----AIRNLREDIREVPKC-ECKKYWKCAIGHN   371
```

QY DTRACD--IIGGLPVASRRGR-----ETILSPANFEGQGRLL-----LAPITAY 206  
Db 372 EAKKCDENWVNSGNNIECESAQSTEDCLAKIYKGEADMSLDGGFFIYIAKCGIVPLAE 431  
QY 207 SQQTRGLLCITLSTLTGDRKQVEGEVQVVSATQSFATCVNGVQWTFPHGASKTLTAG 266  
Db 432 NYETRGASACVDPTEEGYH-----AVAVYKSSSPDILT-----NLSKG 470  
QY 267 PKCFITQWTVNTDQDLVNGQAPFGARSTPCTCGSSDLVYTRHADVLPVRRGDSNGSL 326  
Db 471 KR---SCTGTGADR-TAGNNIPMGJ-----LYSTIKCEPDKFPRBCAAGYR 513  
QY 327 LSPRPVSYLKSGSGGP-LLC-PSGHA-----VGIFRAAVCTRGYAKAVDFIPVSMSE 376  
Db 514 RNSVLNCLIGASAGRGRCPEPNHRIYYGTGAFRCIVKGDVA-----PVKHQIVE 566

Search completed: May 6, 2004, 09:31:52  
Job time : 9.93319 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 35.5407 seconds  
(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-12  
Perfect score: 1987  
Sequence: 1 ALLTSPYKVLARLWML.....RGAAKAVDPFVPSMETTR 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1921	96.7	3010	12 Q9J3F9	Q9J3F9 hepatitis C
2	1918	96.5	3010	12 Q9J3F6	Q9J3F6 hepatitis C
3	1915	96.4	3010	12 Q9J3H7	Q9J3H7 hepatitis C
4	1909	96.1	3010	12 Q9J3D6	Q9J3D6 hepatitis C
5	1908	96.0	3010	12 Q9J3E4	Q9J3E4 hepatitis C
6	1908	96.0	3010	12 Q807P3	Q807P3 hepatitis C
7	1907	96.0	3010	12 Q9J3G6	Q9J3G6 hepatitis C
8	1905	95.9	3010	12 Q68788	Q68788 hepatitis C
9	1905	95.9	3010	12 Q9J3H3	Q9J3H3 hepatitis C
10	1905	95.9	3010	12 Q9J3D7	Q9J3D7 hepatitis C
11	1903	95.8	1186	12 Q81755	Q81755 hepatitis C
12	1903	95.8	1284	12 Q81817	Q81817 hepatitis C
13	1903	95.8	3010	12 P89966	P89966 hepatitis C
14	1902	95.7	3010	12 Q9J3H9	Q9J3H9 hepatitis C
15	1902	95.7	3010	12 Q9J3A2	Q9J3A2 hepatitis C
16	1902	95.7	3010	12 Q9J3I0	Q9J3I0 hepatitis C

17	1902	95.7	3010	12 Q9QIX6	Q9QIX6 hepatitis C
18	1902	95.7	3010	12 Q9QIX5	Q9QIX5 hepatitis C
19	1901	95.7	3008	12 Q9J3F4	Q9J3F4 hepatitis C
20	1900	95.6	3010	12 Q9QIX3	Q9QIX3 hepatitis C
21	1900	95.6	3010	12 P90191	P90191 hepatitis C
22	1899	95.6	3010	12 Q9J3H0	Q9J3H0 hepatitis C
23	1898	95.5	3010	12 Q81760	Q81760 hepatitis C
24	1898	95.5	3010	12 Q9QIX8	Q9QIX8 hepatitis C
25	1898	95.5	3010	12 Q9QIX7	Q9QIX7 hepatitis C
26	1896	95.4	3010	12 Q9J3I1	Q9J3I1 hepatitis C
27	1896	95.4	3010	12 Q9QIX5	Q9QIX5 hepatitis C
28	1895	95.4	3010	12 P88603	P88603 hepatitis C
29	1895	95.4	3010	12 Q9J3H5	Q9J3H5 hepatitis C
30	1895	95.4	3014	12 Q9DTE9	Q9DTE9 hepatitis C
31	1893	95.3	3010	12 Q9DTE9	Q9DTE9 hepatitis C
32	1892	95.2	3010	12 Q68826	Q68826 hepatitis C
33	1892	95.2	3010	12 Q9J3H6	Q9J3H6 hepatitis C
34	1892	95.2	3010	12 Q9DTE9	Q9DTE9 hepatitis C
35	1891	95.2	3010	12 Q9DTE3	Q9DTE3 hepatitis C
36	1889	95.1	3011	12 Q9DTE3	Q9DTE3 hepatitis C
37	1889	95.1	3015	12 Q9WPH5	Q9WPH5 hepatitis C
38	1887	95.0	3010	12 Q9WPH5	Q9WPH5 hepatitis C
39	1885	94.9	3010	12 Q02829	Q02829 hepatitis C
40	1885	94.9	3010	12 Q9QIX6	Q9QIX6 hepatitis C
41	1885	94.9	3010	12 Q9J786	Q9J786 hepatitis C
42	1884	94.8	3010	12 Q9DTE5	Q9DTE5 hepatitis C
43	1882	94.7	3010	12 Q9J3G3	Q9J3G3 hepatitis C
44	1881	94.7	1275	12 Q06642	Q06642 hepatitis C
45	1881	94.7	3010	12 Q9WIK8	Q9WIK8 hepatitis C

#### ALIGNMENTS

##### RESULT 1

Q9J3F9 PRELIMINARY; PRT: 3010 AA.

AC Q9J3F9; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitis C virus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.

RA STRAIN=MD3;  
RC Nagayama K., Kuroaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease progression."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).

DR EMBL AF207774.1, AF65964.1, --  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.  
DR PIR; P50329; P50329.  
DR HSSP; P27958; 1HE1.  
DR MEROPS; S29.001; --  
DR MEROPS; U39.001; --  
DR GO; GO:0016021; C: integral to membrane; IEA.  
DR GO; GO:0019028; C: viral capsid; IEA.  
DR GO; GO:0019031; C: viral envelope; IEA.  
DR GO; GO:0005524; F: ATP binding; IEA.  
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F: electron transporter activity; IEA.  
DR GO; GO:0003723; F: RNA binding; IEA.  
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR000345; CysC\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4p.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PsVlr.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01506; HCV\_NS4p; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KM Coat protein: Envelope protein: Glycoprotein: Nonstructural protein:  
 KM Polyprotein: RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9DB93B5E0C7 CRC64;

Query Match 96.7%; Score 1921; DB 12; Length 3010;  
 Best Local Similarity 95.0%; Pred. No. 5.5e-153;  
 Matches 36; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 ALTLSPYKVLARLILWQYITRVEALQVIRPILNVRGRDAIILLTCANPELIF 60  
 DB 827 ALTLSPYKVFARLILWQYITRVEALHVVWVPLNVRGRDAIILLTCANPELIF 886  
 QY 61 DITKLALIFGLPMLVQAGITKVPYVRAQGLIRACMLVRKXAGHYVQMAEMKLAALTG 120  
 DB 887 DITKLALIFGLPMLVQAGITRPFYVRAQGLIRACMLVRKXAGHYVQMAEMKLAALTG 946  
 QY 121 TVYIDHLTPLOMAHAGLADLAVAVPVTFSMEVKIITWGDTPAACGDIISGLPVSAR 180  
 DB 947 TVYIDHLTPLOMAHAGLADLAVAVPVTFSMETKIITWGDTPAACGDIISGLPVSAR 1006  
 QY 181 GREIILGPADNFGGQWRLLAPITAYSQOTRGLIGIITSLGRDKNQVEGVQVVSAT 240  
 DB 1007 GREIILGPADNFGGQWRLLAPITAYSQOTRGLIGIITSLGRDKNQVEGVQVVSAT 1066  
 QY 241 QSFATLVGVGCVTFVHAGSKTLAAGPKPIITQMTYVNDQDVLVQAPPGASMTPTCG 300  
 DB 1067 QSFATLVGVGCVTFVHAGSKTLAAGPKPIITQMTYVNDQDVLVQAPPGASMTPTCG 1126  
 QY 301 SSGLVYVTHADVTPVRRGDSGSLSPRPVSYLKSGSGGULLCPGSHAVGIFPAVCT 360  
 DB 1127 SSGLVYVTHADVTPVRRGDSGSLSPRPVSYLKSGSGGULLCPGSHAVGIFPAVCT 1186  
 QY 361 RGVAKAVDFIPVESMETTW 380  
 DB 1187 RGVAKAVDFIPVESMETTW 1206

RESULT 2  
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 ID Q9DTE6  
 AC Q9DTE6  
 DT 01-VAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-VAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OK NCBI\_taxid=11103;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HCV1142;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Mario H., Baba K., Hijioka M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 with hepatocellular carcinoma: the 'progression score' revisited.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1-SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AB049091; BAB18804.1; .  
 DR PIR; A61196; A61196.  
 DR PIR; PS0329; PS0329.  
 DR HSSP; P26663; LUXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003958; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4p.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PsVlr.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01506; HCV\_NS4p; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.





Db 1127 SSDLVLTVRHADVIPIVRRRDSRGSLSLSPRVSYLKSSGGPLLCPSGNAVGIFRAAVCT 1186  
 QY 361 GVAKAVDFIPVESMETTR 380  
 Db 1187 GVAKAVDFIPVESMETTR 1206

RESULT 4  
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 ID Q9DTE6  
 AC Q9DTE6; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCVt221;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Marto H., Baba K., Hijioka M.,  
 RA Mishiuro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AB049101; BAB1814.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: PS0329; PS0329.  
 DR HSP: P26653; LUXP.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0016787; F: hydrolase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F: transcription; IEA.  
 DR GO: GO:0019079; F: viral genome replication; IEA.  
 DR GO: GO:0019087; F: viral genome replication; IEA.  
 DR InterPro: IPR009003; Cys Ser typsin.  
 DR InterPro: IPR000345; CysC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD\_heme\_BS.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid\_1.  
 DR Pfam: PF01542; HCV\_core\_1.  
 DR Pfam: PF01539; HCV\_env\_1.

DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PF0186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR ATP-binding; Coat protein; Glycoprotein; Helicase;  
 KW Hydrolase; Nonstructural protein; Polypeptide;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327108 MW; DB182D8103F78B4 CRC64;

Query Match 96.1%; Score 1909; DB 12; Length 3010;  
 Best Local Similarity 95.3%; Pred. No. 5,7e-152;  
 Matches 361; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 2 LLTSPYKVLARLIMWLOVITRVEAHLOVWTPINVRGSRDAIILTCVHPELIFD 61  
 Db 828 LLTSPYKVLARLIMWLOVITRVEAHLOVWTPINVRGSRDAIILTCVHPELIFD 887  
 QY 62 ITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHVQVAFMKLAALVGT 121  
 Db 888 ITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHVQVAFMKLAALVGT 947  
 QY 122 VYVHPLTLPDMNAGRLDLAVNEPIESDMERKITTWADPAACGDIISGLPVARRG 181  
 Db 948 VYVHPLTLPDMNAGRLDLAVNEPIESDMERKITTWADPAACGDIISGLPVARRG 1007  
 QY 182 REILLAGPADNFEQGMWLLAPITAYSOOTRGLGCIITSLTGRDKNOVEGEVOSTATQ 241  
 Db 1008 REILLAGPADNFEQGMWLLAPITAYSOOTRGLGCIITSLTGRDKNOVEGEVOSTATQ 1067  
 QY 242 SFLATCVNGVCMYFTHAGSKTLAGEPKGPTTQMTYTNVDQVLVQKAPPAKRSMTPTCTGS 301  
 Db 1068 SFLATCVNGVCMYFTHAGSKTLAGEPKGPTTQMTYTNVDQVLVQKAPPAKRSMTPTCTGS 1127  
 QY 302 SDDLTVTRHADVIPIVRRRDSRGSLSLSPRVSYLKSSGGPLLCPSGNAVGIFRAAVCTR 361  
 Db 1128 SDDLTVTRHADVIPIVRRRDSRGSLSLSPRVSYLKSSGGPLLCPSGNAVGIFRAAVCTR 1187  
 QY 362 GVAKAVDFIPVESMETTR 380  
 Db 1188 GVAKAVDFIPVESMETTR 1206

RESULT 5  
 Q9DTE4 PRELIMINARY; PRT; 3010 AA.  
 ID Q9DTE4  
 AC Q9DTE4; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCVt150;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohta Y., Kanai K., Marto H., Baba K., Hijioka M.,  
 RA Mishiuro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 RT with hepatocellular carcinoma: the 'progression score' revisited.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

## PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL, AB048093, BAB18806.1, -.

DR PIR, A61196, A61196.

DR PIR, P00246, P00246.

DR PIR, P00804, P00804.

DR PIR, P03029, P03029.

DR HSPB, P26663, 1JXP.

DR GO, GO:0016021, C: integral to membrane, IEA.

DR GO, GO:0018028, C: viral capsid, IEA.

DR GO, GO:0019031, C: viral envelope, IEA.

DR GO, GO:0005524, F: ATP binding, IEA.

DR GO, GO:0008026, F: ATP dependent helicase activity, IEA.

DR GO, GO:0005489, F: electron transporter activity, IEA.

DR GO, GO:0003723, F: RNA binding, IEA.

DR GO, GO:0003968, F: RNA-directed RNA polymerase activity, IEA.

DR GO, GO:0008236, F: serine-type peptidase activity, IEA.

DR GO, GO:0005198, F: structural molecule activity, IEA.

DR GO, GO:0016740, F: transferase activity, IEA.

DR GO, GO:0006118, F: electron transport, IEA.

DR GO, GO:0006508, P: proteolysis and peptidolysis, IEA.

DR GO, GO:0006350, P: transcription, IEA.

DR GO, GO:0018079, P: viral genome replication, IEA.

DR GO, GO:0015087, P: viral transformation, IEA.

DR Interpro, IPR009003, Cys\_Ser\_typsin.

DR Interpro, IPR001403, Cys\_heme\_BS.

DR Interpro, IPR002522, HCV core.

DR Interpro, IPR002521, HCV core.

DR Interpro, IPR002519, HCV env.

DR Interpro, IPR002531, HCV NS1.

DR Interpro, IPR002518, HCV NS2.

DR Interpro, IPR000745, HCV NS4a.

DR Interpro, IPR001490, HCV NS4b.

DR Interpro, IPR002868, HCV NS5a.

DR Interpro, IPR002166, HCV RdRp.

DR Interpro, IPR001650, Helicase\_C.

DR Interpro, IPR004109, Peptidase\_C29.

DR Interpro, IPR007095, RNA\_pol\_DS\_PS.

DR Interpro, IPR007094, RNA\_pol\_PSVir.

DR Pfam, PF01543, HCV core, 1.

DR Pfam, PF01542, HCV core, 1.

DR Pfam, PF01539, HCV env, 1.

DR Pfam, PF01560, HCV NS1, 1.

DR Pfam, PF01538, HCV NS2, 1.

DR Pfam, PF02907, HCV NS3, 1.

DR Pfam, PF01006, HCV NS4a, 1.

DR Pfam, PF01001, HCV NS4b, 1.

DR Pfam, PF01506, HCV NS5a, 1.

DR Pfam, PF00271, helicase\_C, 1.

DR Pfam, PF00998, Viral\_RdRp, 1.

DR ProDom, PD186062, HCV\_NS1, 1.

DR SMART, SM00487, DEXDC, 1.

DR PROSITE, PS00190, CYTOCHROME C, 1.

DR Coat protein, Envelope protein, Glycoprotein, Nonstructural protein, RNA polymerase, RNA-directed RNA polymerase, Transmembrane, Polyprotein, RNA-directed RNA polymerase, Transmembrane.

SEQUENCE 3010 AA, 327324 MW, 3DBCF249BD1151C CMC64.

Query Match 96.08; Score 1908; DB 12; Length 3010;  
Best Local Similarity 94.58; Pred. No. 6.9e-152;  
Matches 359; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 1 ALLTSLSYVYVLLARLIMWCYILTRVEAHLOQVWIPPLNRRGGDAIILITCAVHPELLF 60  
DB 827 ALLTSLSYVYVLLARLIMWCYILTRVEAHLOQVWIPPLNRRGGDAIILITCAVHPELLF 886

QY 61 DITKLLLAIRGLPLNVAAGITKVPYFVRAOGLIRACMLVKAAGHYVQAEMKLAALTG 120  
DB 887 DITKLLLAIRGLPLNVAAGITKVPYFVRAOGLIRACMLVKAAGHYVQAEMKLAALTG 946

QY 121 TYVVDHLTPLODMAHAGRIDLAIVEPVIISDMVKIITGADPAACGDIISGLPVSARR 180  
DB 947 TYVVDHLTPLODMAHAGRIDLAIVEPVIISDMVKIITGADPAACGDIISGLPVSARR 1006

QY 181 GREILLSPADNFEQCGNELLAPITTAISQOTRGLICLIISLTGRDNQVGEVQVYSTAT 240  
DB 1007 GREILLSPADNFEQCGNELLAPITTAISQOTRGLICLIISLTGRDNQVGEVQVYSTAT 1066

QY 241 QSFPLATGVNVCMTVPFAGASKTLAPKGPITQMTYTNVDDLVGMQAPFGARSMTCTCG 300  
DB 1067 QSFPLATGVNVCMTVPFAGASKTLAPKGPITQMTYTNVDDLVGMQAPFGARSMTCTCG 1126

QY 301 SSDLYLTRHADYIPVRRDSSGSLISPPPVSYLKGSSGGLIPCPSGHAAGIFRAVCT 360  
DB 1127 SSDLYLTRHADYIPVRRDSSGSLISPPPVSYLKGSSGGLIPCPSGHAAGIFRAVCT 1186

QY 361 RGAKAVDFIPVESMETMR 380  
DB 1187 RGAKAVDFIPVESMETMR 1206

RESULT 6  
ID 0807P3 PRELIMINARY; PRT; 3010 AA.  
AC 0807P3;  
DT 01-JUN-2003 (TReMBLrel, 24, Created)  
DT 01-JUN-2003 (TReMBLrel, 24, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel, 25, Last annotation update)  
DE Polypeptide.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MILE;  
RX MEDLINE=22047193; PubMed=12051758;  
RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,  
RA Nio Y., Hosaka M., Miyamori Y., Shimotohno K.,  
RT "Subgenomic replicon derived from a cell line infected with the  
RT hepatitis C virus."  
RT Biochem. Biophys. Res. Commun. 293:993-999(2002).  
DR EMBL, AB080299, BAC54896.1, -.  
DR GO, GO:0019028, C: viral capsid, IEA.  
DR GO, GO:0019031, C: viral envelope, IEA.  
DR GO, GO:0005524, F: ATP binding, IEA.  
DR GO, GO:0008026, F: ATP dependent helicase activity, IEA.  
DR GO, GO:0005489, F: electron transporter activity, IEA.  
DR GO, GO:0003723, F: RNA binding, IEA.  
DR GO, GO:0003968, F: RNA-directed RNA polymerase activity, IEA.  
DR GO, GO:0008236, F: serine-type peptidase activity, IEA.  
DR GO, GO:0005198, F: structural molecule activity, IEA.  
DR GO, GO:0006118, F: electron transport, IEA.  
DR GO, GO:0006508, P: proteolysis and peptidolysis, IEA.  
DR GO, GO:0006350, P: transcription, IEA.  
DR GO, GO:0018079, P: viral genome replication, IEA.  
DR GO, GO:0018087, P: viral transformation, IEA.  
DR Interpro, IPR009003, Cys\_Ser\_typsin.  
DR Interpro, IPR000345, Cys\_heme\_BS.  
DR Interpro, IPR001410, DEAD.  
DR Interpro, IPR002522, HCV capsid.  
DR Interpro, IPR002521, HCV core.  
DR Interpro, IPR002519, HCV env.  
DR Interpro, IPR002531, HCV NS1.  
DR Interpro, IPR002518, HCV NS2.  
DR Interpro, IPR000745, HCV NS4a.  
DR Interpro, IPR001490, HCV NS4b.  
DR Interpro, IPR002868, HCV NS5a.  
DR Interpro, IPR002166, HCV RdRp.  
DR Interpro, IPR001650, Helicase\_C.  
DR Interpro, IPR004109, Peptidase\_C29.  
DR Interpro, IPR007095, RNA\_pol\_DS\_PS.  
DR Interpro, IPR007094, RNA\_pol\_PSVir.  
DR Pfam, PF01543, HCV capsid, 1.  
DR Pfam, PF01542, HCV core, 1.  
DR Pfam, PF01539, HCV env, 1.  
DR Pfam, PF01560, HCV NS1, 1.

DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR Pfam; PD186062; HCV\_NS1; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
KW Polypeptidase.  
SQ SEQUENCE 3010 AA; 327097 MW; E86418C7A723E686 CRC64;

Query Match 96.0%; Score 1908; DB 12; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 6,9e-152;  
Matches 360; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 ALTLSPYKVVLLARLIMVQYITREVAHLQVMPPLNVRGGRDAIILLTCAVPELIF 60  
DB 827 AFLTSPYKVVFLARLIMVQYITREVAHLQVMPPLNVRGGRDAIILLTCAVPELIF 886  
QY 61 DITKLLAIFGPMVQAGITRKYFVRAGGIRACMLVRKKAAGHYVQMAFVKLAALNG 120  
DB 887 DITKLLAIFGPMVQAGITRKYFVRAGGIRACMLVRKKAAGHYVQMAFVKLAALNG 946  
QY 121 TYYVDHLTPLOMAHAGLRLDAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASAR 180  
DB 947 TYYVDHLTPLOMAHAGLRLDAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASAR 1006  
QY 181 GRELLGPDNFEQGMRLAPITAYSOQTRGLGIIISLTGRDNQVEGVEVQVYSTAT 240  
DB 1007 GRELLGPDNFEQGMRLAPITAYSOQTRGLGIIISLTGRDNQVEGVEVQVYSTAT 1066  
QY 241 QSEFLATGVNVCWTFVHGAGSKTLAPKPGITQMTYVNDQDVLVGMQAPFAGARSMTPTCG 300  
DB 1067 QSEFLATGVNVCWTFVHGAGSKTLAPKPGITQMTYVNDQDVLVGMQAPFAGARSMTPTCG 1126  
QY 301 SSRLVLTTRHADYTPRRRGDSGSLISPRPVSYLKSGSGGRLCPGSHAVIIFPAVCT 360  
DB 1127 SSRLVLTTRHADYTPRRRGDSGSLISPRPVSYLKSGSGGRLCPGSHAVIIFPAVCT 1186  
QY 361 RGAKAVDFIPVSMETTKR 380  
DB 1187 RGAKAVDFIPVSMETTKR 1206

RESULT 7  
Q9J3G6 PRELIMINARY; PRT; 3010 AA.  
AC Q9J3G6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
OS Genome polyprotein.  
DE Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitis C virus.  
OX NCBI\_Taxid=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=MD26;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease  
progression."  
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBMITTER: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND MRNA (BY SIMILARITY).  
CC EMBL; AF207767; AAF65957.1; -  
DR PIR; A61196; A61196.  
DR PIR; P00246; P00246.

DR PIR; P00254; P00254.  
DR PIR; PS0329; PS0329.  
DR HSB; P26663; 10XP.  
DR MEROPS; S29.002; -  
DR GO; GO:0016021; C: integral to membrane; IEA.  
DR GO; GO:0019028; C: viral capsid; IEA.  
DR GO; GO:0019031; C: viral envelope; IEA.  
DR GO; GO:0005524; F: ATP binding; IEA.  
DR GO; GO:0008026; F: AMP dependent helicase activity; IEA.  
DR GO; GO:0005489; F: electron transporter activity; IEA.  
DR GO; GO:0003723; F: RNA binding; IEA.  
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F: structural molecule activity; IEA.  
DR GO; GO:0016740; F: transferase activity; IEA.  
DR GO; GO:0006118; F: electron transport; IEA.  
DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P: transcription; IEA.  
DR GO; GO:0019079; P: viral genome replication; IEA.  
DR GO; GO:0019087; P: viral transformation; IEA.  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DRAD.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRp.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA pol DS\_PS.  
DR InterPro; IPR007094; RNA pol PSvir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR PRODOM; PD186062; HCV\_NS1; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polypeptidase; RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ SEQUENCE 3010 AA; 327165 MW; 74FAB6B80F24837B CRC64;

Query Match 96.0%; Score 1907; DB 12; Length 3010;  
Best Local Similarity 95.0%; Pred. No. 8,4e-152;  
Matches 361; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 ALTLSPYKVVLLARLIMVQYITREVAHLQVMPPLNVRGGRDAIILLTCAVPELIF 60  
DB 827 AFLTSPYKVVFLARLIMVQYITREVAHLQVMPPLNVRGGRDAIILLTCAVPELIF 886  
QY 61 DITKLLAIFGPMVQAGITRKYFVRAGGIRACMLVRKKAAGHYVQMAFVKLAALNG 120  
DB 887 DITKLLAIFGPMVQAGITRKYFVRAGGIRACMLVRKKAAGHYVQMAFVKLAALNG 946  
QY 121 TYYVDHLTPLOMAHAGLRLDAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASAR 180  
DB 947 TYYVDHLTPLOMAHAGLRLDAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASAR 1006  
QY 181 GRELLGPDNFEQGMRLAPITAYSOQTRGLGIIISLTGRDNQVEGVEVQVYSTAT 240  
DB 1007 GRELLGPDNFEQGMRLAPITAYSOQTRGLGIIISLTGRDNQVEGVEVQVYSTAT 1066

QY 241 QSEIATCVAGVCMVTHGAGSKTLAGPKPITOMYTNVDQDLVGNQAPPGASMPCTCG 300  
 DB 1067 QSEIATCVAGVCMVTHGAGSKTLAGPKPITOMYTNVDQDLVGNQAPPGASMPCTCG 1126  
 QY 301 SSDLYLVTRHADYIPVRRRGRSGSLSPRPVSYLKSGSGGPGLLCPGSHAVGIFRAAVCT 360  
 DB 1127 SSDLYLVTRHADYIPVRRRGRSGSLSPRPISYLGSSGGPGLLCPGSHAVGIFRAAVCT 1186  
 QY 361 RGVAKAVDPVPSMETTMR 380  
 DB 1187 RGVAKAVDPVPSMETTMR 1206

RESULT 8  
 ID Q68788 PRELIMINARY; PRT; 3010 AA.  
 AC Q68788;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE HCV polyprotein (Genome polyprotein).  
 OS Hepatitis C virus  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96362158; PubMed=8720135;  
 RA Seki M., Honda Y.;  
 RT Phosphorolactate antisense oligodeoxynucleotides capable of  
 RT inhibiting Hepatitis C virus gene expression: in vitro translation  
 RT assay.";  
 RL J. Biochem. 118:1199-1204 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; DA5172; BAA08120.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; L1XP.  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003668; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_tyrpsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_NS5a.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C2.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326880 MW; EED840E6A050E766 CRC64;

Query Match 95.9%; Score 1905; DB 12; Length 3010;  
 Best Local Similarity 94.5%; Pred. No. 1.2e-151;  
 Matches 358; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 2 LTTSPPYKVLARIIMWLYLITREAHQWIPPLNVRGGRDAIILITCAVHPELIIFD 61  
 DB 828 LTTSPPYKVLAKLIIMWLYLITREAHQWIPPLNVRGGRDAIILITCAVHPELIIFD 887  
 QY 62 ITKLILAFGLMVLQAGITVPYVRAQGLIRACMLVKAAGHYVMATKLAALITGT 121  
 DB 888 ITKLILAFGLMVLQAGITVPYVRAQGLIRACMLVKAAGHYVMATKLAALITGT 947  
 QY 122 VYVDHLTPLOMAHAGRLDAVAEVPVIFSDEVKIITMGADTAACGDIISGLPVASARG 181  
 DB 948 YIYDHLTPLRMAHAGRLDAVAEVPVIFSDEVKIITMGADTAACGDIISGLPVASARG 1007  
 QY 182 REILIGPDNFEQGCMRLARITAYSQTRGLGCIITSLTGRPKNVGEGVQVASTAQ 241  
 DB 1008 KEILIGPDNFEQGCMRLARITAYSQTRGLGCIITSLTGRPKNVGEGVQVASTAQ 1067  
 QY 242 SFIATCVAGVCMVTHGAGSKTLAGPKPITOMYTNVDQDLVGNQAPPGASMPCTCGS 301  
 DB 1068 SFIATCVAGVCMVTHGAGSKTLAGPKPITOMYTNVDQDLVGNQAPPGASMPCTCGS 1127  
 QY 302 SDLYLVTRHADYIPVRRRGRSGSLSPRPVSYLKSGSGGPGLLCPGSHAVGIFRAAVCTR 361  
 DB 1128 SDLYLVTRHADYIPVRRRGRSGSLSPRPISYLGSSGGPGLLCPGSHAVGIFRAAVCTR 1187  
 QY 362 GYAKAVDPVPSMETTMR 380  
 DB 1188 GYAKAVDPVPSMETTMR 1206

RESULT 9  
 ID Q93JH3 PRELIMINARY; PRT; 3010 AA.  
 AC Q93JH3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD19;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 RT progression.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; AF207760; AAF65950.1; -.

```

DR PIR; A61196; A61196.
DR HSP; PS0329; PS0329.
DR HSP; P26663; 1JXP.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F: electron transporter activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR GO; GO:0016740; F: transferase activity; IEA.
DR GO; GO:0006118; P: electron transport; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_PS_vlr.
DR InterPro; IPR007094; RNA_pol_PS_vlr.
DR Pfam; PF01543; HCV_capsid_1.
DR Pfam; PF01542; HCV_core_1.
DR Pfam; PF01539; HCV_env_1.
DR Pfam; PF01560; HCV_NS1_1.
DR Pfam; PF01538; HCV_NS2_1.
DR Pfam; PF02907; HCV_NS3_1.
DR Pfam; PF01006; HCV_NS4a_1.
DR Pfam; PF01001; HCV_NS4b_1.
DR Pfam; PF01506; HCV_NS5a_1.
DR Pfam; PF00271; helicase_C_1.
DR Pfam; PF00998; Viral_RdRp_1.
DR Pfam; PF018062; HCV_NS1_1.
DR Pfam; PF00487; DEXDC_1.
DR PROSITE; PS00190; CYTOCHROME_C_1.
DR PROSITE; PS00190; CYTOCHROME_C_1.
DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein.
DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
DR SEQUENCE 3010 AA; 327234 MW; 44C34677649CB8DD CRC64;

Query Match 95.9%; Score 1905; DB 12; Length 3010;
Best Local Similarity 93.7%; Pred. No. 1.2e-151;
Matches 356; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 ALLTSPYKVLARLIMWLQVLRVHNLQVLPVLNVRGGRALILLTCAVPELIF 60
DB 827 ALLTSPYKVLARLIMWLQVLRVHNLQVLPVLNVRGGRALILLTCAVPELIF 886
QY 61 DITKLILAFGLPMLVLAQITKVPFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALNG 120
DB 887 SITKLILAFGLPMLVLAQITKVPFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALNG 946
QY 121 TVVYDHLTLFLQMAHAGLDLVAVEPVIFSDMEVKITITWGDPTAACGDIISGLPVSKRR 180
DB 947 TVVYDHLTLFLQMAHAGLDLVAVEPVIFSDMEVKITITWGDPTAACGDIISGLPVSKRR 1006
QY 181 GREILLGPADNFEQGMRLIAPITAVSQOTRGLLCITSLTGKDNQVEGSEVQVSTAT 240
DB 1007 GREILLGPADNFEQGMRLIAPITAVSQOTRGLLCITSLTGKDNQVEGSEVQVSTAT 1066

```

```

QY 241 QSLATCVNGVCFVTFPHGAGSKTLGAPKPIPTQMTNVDPDLVGMQAPPGARSMTPTCTCG 300
DB 1067 QSLATCVNGVCFVTFPHGAGSKTLGAPKPIPTQMTNVDPDLVGMQAPPGARSMTPTCTCG 1126
QY 301 SSDLYLVTRHADVPVRRRGDSGLSPRPVSTYKSSGGPLICPGHAGVIFRAAVCT 360
DB 1127 SSDLYLVTRHADVPVRRRGDSGLSPRPVSTYKSSGGPLICPGHAGVIFRAAVCT 1186
QY 361 RGVAKAVDFVPESEMTTMR 380
DB 1187 RGVAKAVDFVPESEMTTMR 1206

RESULT 10
ID Q9PDD7 PRELIMINARY; PRT; 3010 AA.
AC Q9PDD7.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Genome polypeptin.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CC NCBI_TaxID=11103;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=HCV217;
RA Hatahara T., Ohka Y., Kanai K., Maruo H., Baba K., Hijioka M.,
RA Michio S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.";
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC
DR EMBL; AB049100; BAB18613.1; -.
DR PIR; A61196; A61196.
DR PIR; PQ0804; PQ0804.
DR HSP; P26663; 1JXP.
DR MEROPS; S29_002; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F: electron transporter activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR GO; GO:0016740; F: transferase activity; IEA.
DR GO; GO:0006118; P: electron transport; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.

```



DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF00271; Helicase C; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KM ATP-binding; Helicase; Hydrolase.  
 FT NON\_TER  
 SQ SEQUENCE 1186 AA; 126280 MW; 34170478BA23729A CRC64;  
 Query Match 95.8%; Score 1903; DB 12; Length 1186;  
 Best Local Similarity 94.7%; Pred. No. 5.4e-152;  
 Matches 359; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 LTLSPYKYLRLRLWLOYLITRVEAHQVWIPPLNRRGRDAIILLTCVHPELLFD 61  
 DB 106 LTLSPYKYLRLRLWLOYLITRVEAHQVWIPPLNRRGRDAIILLTCVHPELLFD 165  
 QY 62 ITKLAIIFGLPLVNLQAGITKVPYFVRAOGLIRACMLVKAGHYVQVAFMKLAALTGT 121  
 DB 166 ITKLAIIFGLPLVNLQAGITKVPYFVRAOGLIRACMLVKAGHYVQVAFMKLAALTGT 225  
 QY 122 YVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASARG 181  
 DB 226 YVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASARG 285  
 QY 182 REILGPAHPFEGQGRRLAPITAVSQOTRGLGCIITSLGRDNKQVGEVQVSTATQ 241  
 DB 286 KEILGPAHPFEGQGRRLAPITAVSQOTRGLGCIITSLGRDNKQVGEVQVSTATQ 345  
 QY 242 SPLATCVNGVCMVTFVAGSKTLGAPKGPITQMTNVDDLVGWAQPPARSMPTCTGSG 301  
 DB 346 SPLATCVNGVCMVTFVAGSKTLGAPKGPITQMTNVDDLVGWAQPPARSMPTCTGSG 405  
 QY 302 SDIYLVTRADYTPVARRDSSGSLSPFVSYLKSGSGGELLCSGHAAGVCFRAVACTR 361  
 DB 406 SDIYLVTRADYTPVARRDSSGSLSPFVSYLKSGSGGELLCSGHAAGVCFRAVACTR 465  
 QY 362 GVAKAVDFIPVESMETTMR 380  
 DB 466 GVAKAVDFIPVESMETTMR 484

RESULT 12  
 Q81817  
 ID 081817 PRELIMINARY; PRT; 2284 AA.  
 AC 081817;  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Polypeptide precursor (Genome polypeptide).  
 OS Hepatitis C virus type 2.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepatitis C virus.  
 OC NCBI\_TaxID=40271;  
 RN 11)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94068464; PubMed=7504283;  
 RA Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,  
 RA Akagi T., Kimura K., Shimotohno K.;  
 RT "Proteolytic processing and membrane association of putative  
 RT nonstructural proteins of hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).  
 RN 12)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94333810; PubMed=8056334;  
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;  
 RT "Identification of the domain required for trans-cleavage activity of  
 RT hepatitis C viral serine proteinase.";  
 RL Gene 145:215-219(1994).  
 RN 13)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95056078; PubMed=7966638;  
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;

RT "Hepatitis C virus polypeptide processing: kinetics and mutagenic  
 RT analysis of serine proteinase-dependent cleavage.";  
 RL J. Virol. 68:8418-8422(1994).  
 RN 14)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95156583; PubMed=7853491;  
 RA Tanji Y., Hijikata M., Sato S., Kaneko T., Shimotohno K.;  
 RT "Hepatitis C virus-encoded nonstructural protein NS4A has versatile  
 RT functions in viral protein processing.";  
 RL J. Virol. 69:1575-1581(1995).  
 DR EMBL; D16435; BAA03905.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00249; P00249.  
 DR HSP; P26663; IXP.  
 DR GO; GO:0019012; C:viral ion; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008226; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004108; peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;  
 KW Signal; Transferase.  
 FT SIGNAL 4 20  
 FT CHAIN 55 83  
 FT CHAIN 21 54  
 FT CHAIN 84 300  
 FT CHAIN 301 931  
 FT CHAIN 932 985  
 FT CHAIN 986 1246  
 FT CHAIN 1247 1693  
 FT CHAIN 1694 2284  
 FT CHAIN NS5B.  
 SQ SEQUENCE 2284 AA; 247213 MW; DC272A1517046337 CRC64;  
 Query Match 95.8%; Score 1903; DB 12; Length 2284;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-151;  
 Matches 359; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 LTLSPYKYLRLRLWLOYLITRVEAHQVWIPPLNRRGRDAIILLTCVHPELLFD 61  
 DB 102 LTLSPYKYLRLRLWLOYLITRVEAHQVWIPPLNRRGRDAIILLTCVHPELLFD 161  
 QY 62 ITKLAIIFGLPLVNLQAGITKVPYFVRAOGLIRACMLVKAGHYVQVAFMKLAALTGT 121  
 DB 162 ITKLAIIFGLPLVNLQAGITKVPYFVRAOGLIRACMLVKAGHYVQVAFMKLAALTGT 221  
 QY 122 YVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASARG 181  
 DB 222 YVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAAAGDIISGLPVASARG 281



QY 182 REILLGADNFEQGWMLLAPITAVSQOTRGLGCIITSLTGDKNQVGEVQVSTATQ 241  
 Db 282 KEILLGADNFEQGWMLLAPITAVSQOTRGLGCIITSLTGDKNQVGEVQVSTATQ 341  
 QY 242 SFLATCVGVCMTVYHAGSKITLAPRGKPTTQMTYVNDQVLGMQAPPGARSMTPTCTGCS 301  
 Db 342 SFLATCVGVCMTVYHAGSKITLAPRGKPTTQMTYVNDQVLGMQAPPGARSMTPTCTGCS 401  
 QY 302 SDLYLVTRHADVIIVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHVGIFRAAVCTR 361  
 Db 402 SDLYLVTRHADVIIVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHVGIFRAAVCTR 461  
 QY 362 GVAKAVDFIPVESMETTMR 380  
 Db 462 GVAKAVDFIPVESMETTMR 480

RESULT 13

P89366 PRELIMINARY; PRT; 3010 AA.  
 AC P89366;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE RNA for polyprotein (Genome polyprotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 NC NCBL\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=type 1b;  
 RA Tanaka T.;  
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=type 1b;  
 RA TANAKA T.;  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL; D89872; BAA14035.1; .  
 DR PIR; A61196; A61196.  
 DR PIR; F00246; F00246.  
 DR PIR; F00804; F00804.  
 DR PIR; P50329; P50329.  
 DR HSSP; P26663; IUXP.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008216; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0016740; F:structural molecule activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; F:transcription; IEA.  
 DR GO; GO:0019079; F:viral genome replication; IEA.  
 DR GO; GO:0019087; F:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_Lysase.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.

DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; Peptidase C29.  
 DR InterPro; IPR007095; RNA pol DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PstIir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01066; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS8; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein: Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327023 MW; E075BD9CFD8D1261 CRC64;

Query Match 95.8%; Score 1903; DB 12; Length 3010;  
 Best local similarity 94.7%; Pred. No. 1.8e-151;  
 Matches 359; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 LLTSPYKYLRLIWMGLYITRYBAHLQVWIPPLVNRGGEDATILTCVHPPLIPD 61  
 Db 828 LLTSPYKYLRLIWMGLYITRYBAHLQVWIPPLVNRGGEDATILTCVHPPLIPD 887  
 QY 62 ITKLLAIIPGLMWLQAGITKVPYFVRAQGLIRACWVRKAGHYVQAFMKALITGT 121  
 Db 888 ITKLLAIIPGLMWLQAGITKVPYFVRAQGLIRACWVRKAGHYVQAFMKALITGT 947  
 QY 122 VYVDHTPLQDMWHAAGRDIAVAPEVIFSDMEVKIITGADTRACGDIISGLPVASRG 181  
 Db 948 VYVDHTPLQDMWHAAGRDIAVAPEVIFSDMEVKIITGADTRACGDIISGLPVASRG 1007  
 QY 182 REILLGADNFEQGWMLLAPITAVSQOTRGLGCIITSLTGDKNQVGEVQVSTATQ 241  
 Db 1008 KEILLGADNFEQGWMLLAPITAVSQOTRGLGCIITSLTGDKNQVGEVQVSTATQ 1067  
 QY 242 SFLATCVGVCMTVYHAGSKITLAPRGKPTTQMTYVNDQVLGMQAPPGARSMTPTCTGCS 301  
 Db 1068 SFLATCVGVCMTVYHAGSKITLAPRGKPTTQMTYVNDQVLGMQAPPGARSMTPTCTGCS 1127  
 QY 302 SDLYLVTRHADVIIVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHVGIFRAAVCTR 361  
 Db 1128 SDLYLVTRHADVIIVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHVGIFRAAVCTR 1187  
 QY 362 GVAKAVDFIPVESMETTMR 380  
 Db 1188 GVAKAVDFIPVESMETTMR 1206

RESULT 14

Q93389 PRELIMINARY; PRT; 3010 AA.  
 AC Q93389;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 NC NCBL\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD13;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;



DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_NS5b.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA\_pol\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_core; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF00998; HCV\_NS5a; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR ProDom; PD186062; HCV\_NS1; 1.  
DR SMART; SM00487; DEXDc; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ SEQUENCE 3010 AA; 327007 MW; 053B9A53B0AB335 CRC64;

Query Match 95.7%; Score 1902; DB 12; Length 3010;  
Best Local Similarity 94.2%; Pred. No. 2.2e-151;  
Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 2 LITLSPYKLLARILIMWLOLITRVEHLOWIIPINVRGGSDAILITCAVHPELIPD 61  
DB 828 LITLSPYKVLARILIMWLOLITRVEHLOWIIPINVRGGSDAILITCAVHPELIPD 887  
QY 62 ITKLALAFGLMVLQAGITVVFVRAQGLIRACMLVRKAGGHVYQNAFMKLAALTGT 121  
DB 888 ITKLALAFGLMVLQAGITVVFVRAQGLIRACMLVRKAGGHVYQNAFMKLAALTGT 947  
QY 122 YYIDHLITLQMAHAGLDLAVAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSAARG 181  
DB 948 YLYNHLITLQMAHAGLDLAVAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSAARG 1007  
QY 182 REILLGPADNFGQGRLLAPITAYSQCTRGILGCIITSLTGRDKNQEGEVQVSTATQ 241  
DB 1008 KEILLGPADNFGQGRLLAPITAYSQCTRGILGCIITSLTGRDKNQEGEVQVSTATQ 1067  
QY 242 SFLLACVNGVCMVTHVHSGSKTLGPKPITQMTTNTVDOLVGMQAPPGARSMTPTCTCGS 301  
DB 1068 SFLLACVNGVCMVTHVHSGSKTLGPKPITQMTTNTVDOLVGMQAPPGARSMTPTCTCGS 1127  
QY 302 SDLYLVTRHADVIPVRRRGDSRGLSLSPRPVSYLKGSSGGPILCPSSGHAAGIFRAAVCTR 361  
DB 1128 SDLYLVTRHADVIPVRRRGDSRGLSLSPRPVSYLKGSSGGPILCPSSGHAAGIFRAAVCTR 1187  
QY 362 GVAKAVDFIPVESMETTMR 380  
DB 1188 GVAKAVDFIPVESMETTMR 1206

Search completed: May 6, 2004, 09:35:47  
Job time : 36.5407 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 14.914 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-12

Perfect score: 1987  
Sequence: 1 ALLTSPYKVLARLIWML.....RGAKAVDFIPVSEMETMR 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfillseq.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1887	95.0	2201	4	US-09-539-601-6
2	1887	95.0	2201	4	US-09-539-601-15
3	1887	95.0	3010	4	US-09-539-601-3
4	1887	95.0	3010	4	US-09-539-601-21
5	1887	95.0	3010	4	US-09-539-601-27
6	1881	94.7	1692	3	US-09-263-933-4
7	1881	94.7	1692	3	US-09-263-933-2
8	1881	94.7	2307	3	US-09-263-933-2
9	1881	94.7	2307	4	US-09-539-601-2
10	1880	94.6	3010	4	US-09-539-601-33
11	1878	94.5	1692	3	US-09-263-933-11
12	1878	94.5	1692	3	US-09-263-933-11
13	1878	94.5	2307	3	US-09-263-933-9
14	1878	94.5	2307	4	US-09-263-933-9
15	1869	94.1	1692	3	US-09-263-933-18
16	1869	94.1	1692	3	US-09-263-933-18
17	1869	94.1	2307	3	US-09-263-933-16
18	1869	94.1	2307	4	US-09-263-933-16
19	1869	94.1	3010	3	US-09-539-601-16
20	1823	91.7	2013	1	US-08-324-977-12
21	1823	91.7	2013	2	US-08-324-977-12
22	1823	91.7	2013	2	US-08-324-977-12
23	1823	91.7	2013	3	US-08-324-977-12
24	1823	91.7	2201	4	US-09-539-601-12
25	1823	91.7	2620	1	US-08-324-977-32
26	1823	91.7	2620	1	US-08-324-977-32
27	1823	91.7	2620	2	US-08-324-977-32

28	1823	91.7	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1823	91.7	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1823	91.7	2621	2	US-08-324-977-36	Sequence 36, Appl
31	1823	91.7	2621	2	US-08-324-977-36	Sequence 36, Appl
32	1823	91.7	2621	3	US-08-324-977-36	Sequence 36, Appl
33	1823	91.7	3010	1	US-08-324-977-14	Sequence 2, Appl
34	1823	91.7	3010	1	US-08-324-977-14	Sequence 2, Appl
35	1823	91.7	3010	2	US-08-324-977-14	Sequence 2, Appl
36	1823	91.7	3010	2	US-08-324-977-14	Sequence 2, Appl
37	1823	91.7	3010	2	US-08-324-977-14	Sequence 2, Appl
38	1823	91.7	3010	2	US-08-324-977-14	Sequence 2, Appl
39	1823	91.7	3010	3	US-08-324-977-14	Sequence 2, Appl
40	1823	91.7	3010	3	US-08-324-977-14	Sequence 2, Appl
41	1718	86.5	1648	1	PCT-US94-07280-12	Sequence 12, Appl
42	1718	86.5	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1718	86.5	3011	1	US-08-188-2818-1	Sequence 1, Appl
44	1718	86.5	3011	1	US-08-188-2818-1	Sequence 1, Appl
45	1718	86.5	3011	1	US-08-453-552-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-539-601-6  
Sequence 6, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartschlagel, Ralf FM  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 2201  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match 95.0% Score 1887; DB 4; Length 2201;  
Best Local Similarity 93.4% Pred. No. 3.8e-181;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY	2	ALLTSPYKVLARLIWMLQYLRVNAHLOWIPLNVRGGRDAIILTCVAHEBELIFD	61
DB	19	LLTSPHYKFLARLIWMLQYLRVNAHLOWIPLNVRGGRDAIILTCVAHEBELIFD	78
QY	62	ITKLLAIFGPIVMAQGITVPRVRAQGLIRACMYKAAAGHYVQMAEMKLAALGT	121
DB	79	ITKLLAIFGPIVMAQGITVPRVRAQGLIRACMYKAAAGHYVQMAEMKLAALGT	138
QY	122	YYVDHITPLQDAHAGLRLDAVAEPVIFSDMEVKIITMGADTAACGDIIISGLPVASARG	181
DB	139	YYVDHITPLQDAHAGLRLDAVAEPVIFSDMEVKIITMGADTAACGDIIISGLPVASARG	198
QY	182	REILGPDNFEQGRRLIAPITANSQOTRGLIGITSLGRDNQVGEVQVSTANTQ	241
DB	199	REIHGPDNFEQGRRLIAPITANSQOTRGLIGITSLGRDNQVGEVQVSTANTQ	258
QY	242	SLATCAVGCWTVFVHGAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTG	301
DB	259	SLATCAVGCWTVFVHGAGSKTLAAGPKPITQMTYTNVDQDLVGMQAPPGASMTPTCTG	318
QY	302	SDLYIVTRHADYIVRRRGRSGSLSPRPVSYLKGSSGCPILCSGAVGIFRAAVCTR	361
DB	319	SDLYIVTRHADYIVRRRGRSGSLSPRPVSYLKGSSGCPILCSGAVGIFRAAVCTR	378
QY	362	GVAKAVDFIPVSEMETMR 380	

Db 379 GAAKAVDFVVESEMETTMR 397

## RESULT 2

US-09-539-601-15  
 ; Sequence 15; Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 15  
 ; LENGTH: 2201  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 ; US-09-539-601-15

Query Match 95.0%; Score 1887; DB 4; Length 2201;  
 Best Local Similarity 93.4%; Pred. No. 3,8e-181;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Db 2 LTLTSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGDAIILLTCAVHPELIFD 61  
 19 LTLTSPHYKFLARLIMWLQYFITRAEHLQVWIPPLNVRGGDAVILLTCAIHPELIFT 78  
 QY 62 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGT 121  
 Db 79 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGT 138  
 QY 122 YVVDHLTPLODMNAGLRDLAVAVEPVIFSDMEVKIITWADTAACGDIISGLPVASARG 181  
 Db 139 YVVDHLTPLODMNAGLRDLAVAVEPVIFSDMEVKIITWADTAACGDIISGLPVASARG 198  
 QY 182 REILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQ 241  
 Db 199 REILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQ 258  
 QY 242 SFATCVNVCMTVPHGAGSKTLGPKGPIITOMYTNVDDLVGMQAPGARSMTPCTCGS 301  
 Db 259 SFATCVNVCMTVPHGAGSKTLGPKGPIITOMYTNVDDLVGMQAPGARSMTPCTCGS 318  
 QY 302 SDLYLVRHADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
 Db 319 SDLYLVRHADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 378  
 QY 362 GAAKAVDFVVESEMETTMR 380  
 Db 379 GAAKAVDFVVESEMETTMR 397

## RESULT 3

US-09-539-601-3  
 ; Sequence 3; Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 3010  
 ; TYPE: PRT

ORGANISM: Hepatitis C virus  
 US-09-539-601-3

Query Match 95.0%; Score 1887; DB 4; Length 3010;  
 Best Local Similarity 93.4%; Pred. No. 6.1e-181;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLTSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGDAIILLTCAVHPELIFD 61  
 Db 828 LTLTSPHYKFLARLIMWLQYFITRAEHLQVWIPPLNVRGGDAVILLTCAIHPELIFT 887  
 QY 62 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGT 121  
 Db 888 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGT 947  
 QY 122 YVVDHLTPLODMNAGLRDLAVAVEPVIFSDMEVKIITWADTAACGDIISGLPVASARG 181  
 Db 948 YVVDHLTPLODMNAGLRDLAVAVEPVIFSDMEVKIITWADTAACGDIISGLPVASARG 1007  
 QY 182 REILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQ 241  
 Db 1008 REILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQ 1067  
 QY 242 SFATCVNVCMTVPHGAGSKTLGPKGPIITOMYTNVDDLVGMQAPGARSMTPCTCGS 301  
 Db 1068 SFATCVNVCMTVPHGAGSKTLGPKGPIITOMYTNVDDLVGMQAPGARSMTPCTCGS 1127  
 QY 302 SDLYLVRHADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361  
 Db 1128 SDLYLVRHADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 1187  
 QY 362 GAAKAVDFVVESEMETTMR 380  
 Db 1188 GAAKAVDFVVESEMETTMR 1206

## RESULT 4

US-09-539-601-21  
 ; Sequence 21; Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barteneschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 21  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 ; US-09-539-601-21

Query Match 95.0%; Score 1887; DB 4; Length 3010;  
 Best Local Similarity 93.4%; Pred. No. 6.1e-181;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLTSPYKYLARLIMWLQYLITRVEAHQVWIPPLNVRGGDAIILLTCAVHPELIFD 61  
 Db 828 LTLTSPHYKFLARLIMWLQYFITRAEHLQVWIPPLNVRGGDAVILLTCAIHPELIFT 887  
 QY 62 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGT 121  
 Db 888 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAGHYVQMAFMKLAALTGT 947  
 QY 122 YVVDHLTPLODMNAGLRDLAVAVEPVIFSDMEVKIITWADTAACGDIISGLPVASARG 181  
 Db 948 YVVDHLTPLODMNAGLRDLAVAVEPVIFSDMEVKIITWADTAACGDIISGLPVASARG 1007  
 QY 182 REILGPADNFEQGWRLAPITAVSQOTRGLGCIITSLTGRDNQVEGEVQVSTATQ 241

Fri May 7 13:37:02 2004

us-10-650-585-12.ra1

Page 3

```
Db 1008 REIHGPRDSLEGGQWRLAPITAYSQOTRGLGCIITSLTGRDRNKGEGEVQVSTATQ 1067
Qy 242 SFLATCVNGVCMVTVHGAAGSKTLAPKPGKITOMYTNVQDLVGMQAPPGARSLPCTCGS 301
Db 1068 SFLATCVNGVCMVTVHGAAGSKTLAPKPGKITOMYTNVQDLVGMQAPPGARSLPCTCGS 1127
Qy 302 SDLYVTNRADVIPIRRRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
Db 1128 SDLYVTNRADVIPIRRRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 1187
Qy 362 GVAKAVDFIVESMETTMR 380
Db 1188 GVAKAVDFIVESMETTMR 1206
```

```
RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartsch, Ralf FM
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-27
```

```
Query Match 95.0%; Score 1887; DB 4; Length 3010;
Best Local Similarity 93.4%; Pred. No. 6.1e-181;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
```

```
Qy 2 LITLSPYKVLARLIMWLQYLITRVEAHQWIPPLNVRGGRDAIILLTCAVHPELIFD 61
Db 828 LITLSPYKVLARLIMWLQYLITRVEAHQWIPPLNVRGGRDAIILLTCAVHPELIFD 887
Qy 62 ITKLILAFGLPMLVLAQGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGT 121
Db 888 ITKLILAFGLPMLVLAQGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGT 947
Qy 122 YVVDHLPLODMARAGRLDAVAEPIFSDEMEKITTGADTAACGDIISGLPVSARRG 181
Db 948 YVVDHLPLODMARAGRLDAVAEPIFSDEMEKITTGADTAACGDIISGLPVSARRG 1007
Qy 182 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDRNKGEGEVQVSTATQ 241
Db 1008 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDRNKGEGEVQVSTATQ 1067
Qy 242 SFLATCVNGVCMVTVHGAAGSKTLAPKPGKITOMYTNVQDLVGMQAPPGARSLPCTCGS 301
Db 1068 SFLATCVNGVCMVTVHGAAGSKTLAPKPGKITOMYTNVQDLVGMQAPPGARSLPCTCGS 1127
Qy 302 SDLYVTNRADVIPIRRRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
Db 1128 SDLYVTNRADVIPIRRRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 1187
Qy 362 GVAKAVDFIVESMETTMR 380
Db 1188 GVAKAVDFIVESMETTMR 1206
```

```
RESULT 6
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
```

```
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-4
```

```
Query Match 94.7%; Score 1881; DB 3; Length 1692;
Best Local Similarity 93.1%; Pred. No. 1e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;
```

```
Qy 2 LITLSPYKVLARLIMWLQYLITRVEAHQWIPPLNVRGGRDAIILLTCAVHPELIFD 61
Db 107 LITLSPYKVLARLIMWLQYLITRVEAHQWIPPLNVRGGRDAIILLTCAVHPELIFD 166
Qy 62 ITKLILAFGLPMLVLAQGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGT 121
Db 167 ITKLILAFGLPMLVLAQGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALGT 226
Qy 122 YVVDHLPLODMARAGRLDAVAEPIFSDEMEKITTGADTAACGDIISGLPVSARRG 181
Db 227 YVVDHLPLODMARAGRLDAVAEPIFSDEMEKITTGADTAACGDIISGLPVSARRG 286
Qy 182 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDRNKGEGEVQVSTATQ 241
Db 287 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDRNKGEGEVQVSTATQ 346
Qy 242 SFLATCVNGVCMVTVHGAAGSKTLAPKPGKITOMYTNVQDLVGMQAPPGARSLPCTCGS 301
Db 347 SFLATCVNGVCMVTVHGAAGSKTLAPKPGKITOMYTNVQDLVGMQAPPGARSLPCTCGS 406
Qy 302 SDLYVTNRADVIPIRRRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
Db 407 SDLYVTNRADVIPIRRRGDSRGSLSLSPRVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 466
Qy 362 GVAKAVDFIVESMETTMR 380
Db 467 GVAKAVDFIVESMETTMR 485
```

```
RESULT 7
US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; EARLIER APPLICATION NUMBER: 09/263,933
; EARLIER FILING DATE: 1998-02-08
; PRIOR FILING DATE: 1999-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-4

Query Match      94.7%; Score 1881; DB 4; Length 1692;
Best Local Similarity 93.1%; Pred. No. 1e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLRLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
DB 107 LTLSPYKVLRLRLIMLQYFTTRAEHLHWIIPPLNVRGGRDAIILLTCAVHPELIFD 166
QY 62 ITKLALIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALALGT 121
DB 167 ITKLALIFGPIPLMVLQAGITRVPYFRAOGLIRACMLYKKAAGHYVQMAFKALALGT 226
QY 122 YVVDHLTPLODMWHAAGLRDLAVAVEPVFSDEMKIITWGADTAACGDIISGLPVASARG 181
DB 227 YVVDHLTPLODMWHAAGLRDLAVAVEPVFSDEMKIITWGADTAACGDIISGLPVASARG 286
QY 182 REILGPADNFEQGRMLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 241
DB 287 KEILGPADNFEQGRMLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 346
QY 242 SPLATCVNGVCWTVFHGAGSKTLGPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCTG 301
DB 347 SPLATCVNGVCWTVFHGAGSKTLGPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCTG 406
QY 302 SDLYLTRADVIYVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 361
DB 407 SDLYLTRADVIYVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 466
QY 362 GVAKAVDFIVPESMETTMR 380
DB 467 GVAKAVDFIVPESMETTMR 485

RESULT 8
US-09-263-933-2
; Sequence 2, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1999-03-08
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-2

Query Match      94.7%; Score 1881; DB 3; Length 2307;
Best Local Similarity 93.1%; Pred. No. 1.7e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLRLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
DB 199 LTLSPYKVLRLRLIMLQYFTTRAEHLHWIIPPLNVRGGRDAIILLTCAVHPELIFD 256
QY 62 ITKLALIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALALGT 121
DB 259 ITKLALIFGPIPLMVLQAGITRVPYFRAOGLIRACMLYKKAAGHYVQMAFKALALGT 318
```

```

QY 122 YVVDHLTPLODMWHAAGLRDLAVAVEPVFSDEMKIITWGADTAACGDIISGLPVASARG 181
DB 319 YVVDHLTPLODMWHAAGLRDLAVAVEPVFSDEMKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPADNFEQGRMLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 241
DB 379 KEILGPADNFEQGRMLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 438
QY 242 SPLATCVNGVCWTVFHGAGSKTLGPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCTG 301
DB 439 SPLATCVNGVCWTVFHGAGSKTLGPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCTG 498
QY 302 SDLYLTRADVIYVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 361
DB 499 SDLYLTRADVIYVRRRGRSGSLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 558
QY 362 GVAKAVDFIVPESMETTMR 380
DB 559 GVAKAVDFIVPESMETTMR 577

RESULT 9
US-09-919-901-2
; Sequence 2, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; EARLIER FILING DATE: 2001-08-02
; EARLIER FILING DATE: 1999-02-08
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-2

Query Match      94.7%; Score 1881; DB 4; Length 2307;
Best Local Similarity 93.1%; Pred. No. 1.7e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLRLRLIMLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
DB 199 LTLSPYKVLRLRLIMLQYFTTRAEHLHWIIPPLNVRGGRDAIILLTCAVHPELIFD 258
QY 62 ITKLALIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALALGT 121
DB 259 ITKLALIFGPIPLMVLQAGITRVPYFRAOGLIRACMLYKKAAGHYVQMAFKALALGT 318
QY 122 YVVDHLTPLODMWHAAGLRDLAVAVEPVFSDEMKIITWGADTAACGDIISGLPVASARG 181
DB 319 YVVDHLTPLODMWHAAGLRDLAVAVEPVFSDEMKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPADNFEQGRMLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 241
DB 379 KEILGPADNFEQGRMLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 438
QY 242 SPLATCVNGVCWTVFHGAGSKTLGPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCTG 301
DB 439 SPLATCVNGVCWTVFHGAGSKTLGPKGPITQWYTNVDQDLVGMQAPPGARSMTPTCTG 498
```



QY 302 SGLVTRHADVIPIVRRRDSRGSLSPPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
 |||||  
 DB 499 SGLVTRHADVIPIVRRRDSRGSLSPPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 558  
 |||||  
 QY 362 GVAKAVDFPVESEMETTR 380  
 |||||  
 DB 559 GVAKAVDFPVESEMETTR 577  
 |||||

RESULT 10  
 US-09-539-601-33  
 ; Sequence 33, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlager, Ralf FM  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 33  
 ; LENGTH: 3010  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-33

Query Match 94.6%; Score 1880; DB 4; Length 3010;  
 Best Local Similarity 93.1%; Pred. No. 3.1e-180;  
 Matches 353; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMWLOYLTRVEAHQVWIPPLNVRGSDAIIITCAVHPELIIFD 61  
 |||||  
 DB 828 LTLSPYKVLARLIMWLOYLTRVEAHQVWIPPLNVRGSDAIIITCAVHPELIIFD 887  
 |||||  
 QY 62 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121  
 |||||  
 DB 888 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 947  
 |||||  
 QY 122 YVVDHLPLQDMAHAGRLDAVAEPIFSDMEVKIITMGADTAACGDIISGLPVSARG 181  
 |||||  
 DB 948 YVVDHLPLQDMAHAGRLDAVAEPIFSDMEVKIITMGADTAACGDIISGLPVSARG 1007  
 |||||  
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATQ 241  
 |||||  
 DB 1008 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATQ 1067  
 |||||  
 QY 242 SFLATCVNGVCMVYFHGAGSKTLAGPKGPIITQMTNTVDQDVLVGMQAPGARSLPCTCGS 301  
 |||||  
 DB 1068 SFLATCVNGVCMVYFHGAGSKTLAGPKGPIITQMTNTVDQDVLVGMQAPGARSLPCTCGS 1127  
 |||||  
 QY 302 SDLYLVRHADVIPIVRRRDSRGSLSPPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
 |||||  
 DB 1128 SDLYLVRHADVIPIVRRRDSRGSLSPPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 1187  
 |||||  
 QY 362 GVAKAVDFPVESEMETTR 380  
 |||||  
 DB 1188 GVAKAVDFPVESEMETTR 1206  
 |||||

RESULT 11  
 US-09-263-933-11  
 ; Sequence 11, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A

CURRENT APPLICATION NUMBER: US/09/263,933  
 ; CURRENT FILING DATE: 1999-03-08  
 ; EARLIER APPLICATION NUMBER: 09/129,611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 US-09-263-933-11

Query Match 94.5%; Score 1878; DB 3; Length 1692;  
 Best Local Similarity 92.9%; Pred. No. 2.1e-180;  
 Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMWLOYLTRVEAHQVWIPPLNVRGSDAIIITCAVHPELIIFD 61  
 |||||  
 DB 107 LTLSPYKVLARLIMWLOYLTRVEAHQVWIPPLNVRGSDAIIITCAVHPELIIFD 166  
 |||||  
 QY 62 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121  
 |||||  
 DB 167 ITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 226  
 |||||  
 QY 122 YVVDHLPLQDMAHAGRLDAVAEPIFSDMEVKIITMGADTAACGDIISGLPVSARG 181  
 |||||  
 DB 227 YVVDHLPLQDMAHAGRLDAVAEPIFSDMEVKIITMGADTAACGDIISGLPVSARG 286  
 |||||  
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATQ 241  
 |||||  
 DB 287 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDNQVEGEVQVSTATQ 346  
 |||||  
 QY 242 SFLATCVNGVCMVYFHGAGSKTLAGPKGPIITQMTNTVDQDVLVGMQAPGARSLPCTCGS 301  
 |||||  
 DB 347 SFLATCVNGVCMVYFHGAGSKTLAGPKGPIITQMTNTVDQDVLVGMQAPGARSLPCTCGS 406  
 |||||  
 QY 302 SDLYLVRHADVIPIVRRRDSRGSLSPPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361  
 |||||  
 DB 407 SDLYLVRHADVIPIVRRRDSRGSLSPPRVSYLKSGSGPILCPGSHAVGIFRAAVCTR 466  
 |||||  
 QY 362 GVAKAVDFPVESEMETTR 380  
 |||||  
 DB 467 GVAKAVDFPVESEMETTR 485  
 |||||

RESULT 12  
 US-09-919-901-11  
 ; Sequence 11, Application US/09919901  
 ; Patent No. 6599738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patrick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263,933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 US-09-919-901-11

Query Match 94.5%; Score 1878; DB 4; Length 1692;

Best Local Similarity 92.9%; Pred. No. 2,1e-180;  
Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

```

QY 2 LTLSPYKVLRLIWMLOYLITRVEAHLQVWIPPLNVRGGRDAIILLTCVHPELIFD 61
DB 107 LTLSPYKVLRLIWMLOYLITRVEAHLQVWIPPLNVRGGRDAIILLTCVHPELIFD 166
QY 62 ITKLLAIFGPLMWLOAGITRVYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 167 ITKLLAIFGPLMWLOAGITRVYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 226
QY 122 YVVDHLTPIODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB 227 YVVDHLTPIODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 286
QY 182 REILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241
DB 287 KEILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 346
QY 242 SFPLATCVGCVMTVYHAGSKTLGPKGPITQMTYTNVDDDLVGMQAPPGARSMTPTCTGS 301
DB 347 SFPLATCVGCVMTVYHAGSKTLGPKGPITQMTYTNVDDDLVGMQAPPGARSMTPTCTGS 406
QY 302 SDIYVTRHADVIPIVRRRDSRGLSPRPVSYLKSGSGGPIILCPSGHAAGIFRAAVCTR 361
DB 407 SDIYVTRHADVIPIVRRRDSRGLSPRPVSYLKSGSGGPIILCPSGHAAGIFRAAVCTR 466
QY 362 GVAKAVDFIPVESMETTMR 380
DB 467 GVAKAVDFIPVESMETTMR 485

```

## RESULT 13

US-09-263-933-9  
Sequence 9, Application US/09263933

```

; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

```

Query Match 94.5%; Score 1878; DB 3; Length 2307;  
Best Local Similarity 92.9%; Pred. No. 3,3e-180;  
Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

```

QY 2 LTLSPYKVLRLIWMLOYLITRVEAHLQVWIPPLNVRGGRDAIILLTCVHPELIFD 61
DB 199 LTLSPYKVLRLIWMLOYLITRVEAHLQVWIPPLNVRGGRDAIILLTCVHPELIFD 258
QY 62 ITKLLAIFGPLMWLOAGITRVYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 259 ITKLLAIFGPLMWLOAGITRVYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 318
QY 122 YVVDHLTPIODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB 319 YVVDHLTPIODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241
DB 378 REILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241

```

```

DB 379 KEILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 438
QY 242 SFPLATCVGCVMTVYHAGSKTLGPKGPITQMTYTNVDDDLVGMQAPPGARSMTPTCTGS 301
DB 439 SFPLATCVGCVMTVYHAGSKTLGPKGPITQMTYTNVDDDLVGMQAPPGARSMTPTCTGS 498
QY 302 SDIYVTRHADVIPIVRRRDSRGLSPRPVSYLKSGSGGPIILCPSGHAAGIFRAAVCTR 361
DB 499 SDIYVTRHADVIPIVRRRDSRGLSPRPVSYLKSGSGGPIILCPSGHAAGIFRAAVCTR 558
QY 362 GVAKAVDFIPVESMETTMR 380
DB 559 GVAKAVDFIPVESMETTMR 577

```

## RESULT 14

US-09-919-901-9  
Sequence 9, Application US/09919901

```

; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-9

```

Query Match 94.5%; Score 1878; DB 4; Length 2307;  
Best Local Similarity 92.9%; Pred. No. 3,3e-180;  
Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

```

QY 2 LTLSPYKVLRLIWMLOYLITRVEAHLQVWIPPLNVRGGRDAIILLTCVHPELIFD 61
DB 199 LTLSPYKVLRLIWMLOYLITRVEAHLQVWIPPLNVRGGRDAIILLTCVHPELIFD 258
QY 62 ITKLLAIFGPLMWLOAGITRVYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 259 ITKLLAIFGPLMWLOAGITRVYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 318
QY 122 YVVDHLTPIODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB 319 YVVDHLTPIODMAHAGLRDLAVAVEPIFSDMEVKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241
DB 379 REILGPDNFEQGRRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 438
QY 242 SFPLATCVGCVMTVYHAGSKTLGPKGPITQMTYTNVDDDLVGMQAPPGARSMTPTCTGS 301
DB 439 SFPLATCVGCVMTVYHAGSKTLGPKGPITQMTYTNVDDDLVGMQAPPGARSMTPTCTGS 498
QY 302 SDIYVTRHADVIPIVRRRDSRGLSPRPVSYLKSGSGGPIILCPSGHAAGIFRAAVCTR 361
DB 499 SDIYVTRHADVIPIVRRRDSRGLSPRPVSYLKSGSGGPIILCPSGHAAGIFRAAVCTR 558
QY 362 GVAKAVDFIPVESMETTMR 380
DB 559 GVAKAVDFIPVESMETTMR 577

```

RESULT 15  
US-09-263-933-18  
; Sequence 18, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roderica L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/263, 933  
; CURRENT FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: 09/129, 611  
; EARLIER FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO: 18  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
US-09-263-933-18

Query Match 94.1%; Score 1869; DB 3; Length 1692;

Best Local Similarity 92.6%; Pred. No. 1,7e-179;  
Matches 351; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

```
QY      2 LTTLSPPYKVLRLRLIMWLQYLITRVEAHLQVWIPPLNVRGGDAIILITCAVHPELIID 61
DB      107 LTTLSPPYKVLRLRLIMWLQYLITRVEAHLQVWIPPLNVRGGDAIILITCAVHPELIID 166
QY      62 ITKLLAIIGPIMVYQAGITKTPYFPAQGLRACMLVPRKAGHYVQMAFYKLAALTGT 121
DB      167 ITKLLAIIGPIMVYQAGITKTPYFPAQGLRACMLVPRKAGHYVQMAFYKLAALTGT 226
QY      122 VYVDHLTPLODWAHAGLRDLAVAVEPFIISDMEVKIITWADTAAAGDIISGLPVASARRG 181
DB      227 YTNHILTPRDWAHAGLRDLAVAVEPFIISDMEVKIITWADTAAAGDIISGLPVASARRG 286
QY      182 REILIGPADNFEQGKRLAPITAYSQTRGLIGCIITSLTGRDKXQVGEVQVYSTATQ 241
DB      287 KEILIGPADNFEQGKRLAPITAYSQTRGLIGCIITSLTGRDKXQVGEVQVYSTATQ 346
QY      242 SFPLATCVNGVCMVTFHAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCTGSS 301
DB      347 SFPLATCVNGVCMVTFHAGSKTLAGPKGPIITQWYTNVDODLVGMQAPPGARSMTPTCTGSS 406
QY      302 SDLYIVTRHADVIPIVRRGDSRGSLLSPRVSYLKSSGGPILCPGHAAGIFRAAVCTR 361
DB      407 SDLYIVTRHADVIPIVRRGDSRGSLLSPRVSYLKSSGGPILCPGHAAGIFRAAVCTR 466
QY      362 GVAKAVDFIVPESMETTMR 380
DB      467 GVAKAVDFIVPESMETTMR 485
```

Search completed: May 6, 2004, 09:39:03  
Job time: 15.9144 secs



Fri May 7 13:37:02 2004

us-10-650-585-12.rapb

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 38.8727 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-12  
Perfect score: 1987  
Sequence: 1 ALLTSPYKVLARLIMWL.....RGVAKAVDFIVESNETTR 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/1/pubpaa/PCRT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/1/pubpaa/PCRTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1987	100.0	380	US-10-017-736-12	Sequence 12, Appl
2	1987	100.0	380	US-10-650-585-12	Sequence 12, Appl
3	1987	100.0	393	US-10-017-736-11	Sequence 11, Appl
4	1987	100.0	393	US-10-650-585-11	Sequence 11, Appl
5	1987	100.0	409	US-10-017-736-2	Sequence 2, Appl
6	1987	100.0	409	US-10-650-585-2	Sequence 2, Appl
7	1887	95.0	2201	US-10-029-907-3	Sequence 3, Appl
8	1887	95.0	2201	US-10-309-561-3	Sequence 3, Appl
9	1887	95.0	3010	US-10-467-000-1	Sequence 1, Appl
10	1881	94.7	1692	US-09-919-901-4	Sequence 4, Appl
11	1881	94.7	1692	US-10-191-966-4	Sequence 4, Appl
12	1881	94.7	2307	US-09-919-901-2	Sequence 2, Appl
13	1881	94.7	2307	US-10-191-966-2	Sequence 2, Appl
14	1878	94.5	1692	US-09-919-901-11	Sequence 11, Appl
15	1878	94.5	1692	US-10-191-966-11	Sequence 11, Appl

16	1878	94.5	2307	US-09-919-901-9	Sequence 9, Appl
17	1878	94.5	2307	US-10-191-966-9	Sequence 9, Appl
18	1869	94.1	1692	US-09-919-901-18	Sequence 18, Appl
19	1869	94.1	1692	US-10-191-966-18	Sequence 18, Appl
20	1869	94.1	2307	US-09-919-901-16	Sequence 16, Appl
21	1869	94.1	2307	US-10-191-966-16	Sequence 16, Appl
22	1842	92.7	352	US-10-017-736-13	Sequence 13, Appl
23	1842	92.7	352	US-10-650-585-13	Sequence 13, Appl
24	1823	91.7	2201	US-10-085-476-2	Sequence 2, Appl
25	1823	91.7	341	US-10-017-736-14	Sequence 14, Appl
26	1778	89.5	341	US-10-650-585-14	Sequence 14, Appl
27	1778	89.5	3011	US-09-742-659-4	Sequence 4, Appl
28	1717	86.4	3011	US-09-891-894-3	Sequence 3, Appl
29	1717	86.4	3011	US-10-184-150-3	Sequence 3, Appl
30	1717	86.4	3011	US-10-328-997-3	Sequence 3, Appl
31	1717	86.4	3012	US-09-238-076-2	Sequence 2, Appl
32	1717	86.4	3012	US-09-995-937-2	Sequence 2, Appl
33	1715	86.3	3012	US-09-917-563-2	Sequence 2, Appl
34	1715	86.3	3011	US-09-916-359-2	Sequence 2, Appl
35	1715	86.3	3011	US-10-296-734-406	Sequence 406, Appl
36	1713	86.2	3011	US-09-238-076-20	Sequence 20, Appl
37	1713	86.2	3011	US-09-995-937-20	Sequence 20, Appl
38	1713	86.2	3011	US-09-917-563-20	Sequence 20, Appl
39	1710	86.1	2894	US-09-941-611-23	Sequence 23, Appl
40	1710	86.1	2894	US-10-044-995-23	Sequence 23, Appl
41	1710	86.1	3011	US-09-952-572-9	Sequence 9, Appl
42	1710	86.1	3011	US-09-747-419-20	Sequence 20, Appl
43	1710	86.1	3011	US-10-189-359-14	Sequence 14, Appl
44	1710	86.1	3011	US-10-259-275-20	Sequence 20, Appl
45	1707	85.9	2985	US-10-259-275-40	Sequence 40, Appl

# ALIGNMENTS

RESULT 1  
US-10-017-736-12  
; Sequence 12, Application US/10017736  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 380  
; TYPE: PRP  
; ORGANISM: HCV  
; US-10-017-736-12

Query Match	100.0%	Score 1987	DB 13	Length 380
Best Local Similarity	100.0%	Pred. No. 1.2e+13	Indels 0	Gaps 0
Matches 380	Conservative	0	Mismatches 0	
QY	1	ALLTSPYKVLARLIMWLQYLITREVAHLQWIPPLNVRGGRDAIILLTCVAPPELIF	60	
DB	1	ALLTSPYKVLARLIMWLQYLITREVAHLQWIPPLNVRGGRDAIILLTCVAPPELIF	60	
QY	61	DIITKLALFGLMVLQAGITVPYFVRAQGLIRACMLYRKAAGHYVMATKALATG	120	
DB	61	DIITKLALFGLMVLQAGITVPYFVRAQGLIRACMLYRKAAGHYVMATKALATG	120	
QY	121	TYVYCHLTFLQWVAHAGLDAVAEVPVFSMEVXKIIITWAGDTAACGIIISGLPVSARR	180	
DB	121	TYVYCHLTFLQWVAHAGLDAVAEVPVFSMEVXKIIITWAGDTAACGIIISGLPVSARR	180	
QY	181	GREILGPADNFGQGRLLAPITAVSQTRGLGCIITSLTGRDNQVGEVQVSTNT	240	

Db 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKKQVGEVQVSTAT 240  
QY 241 QSFPLATCVNGCVMTVFHAGSKTLAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300  
Db 241 QSFPLATCVNGCVMTVFHAGSKTLAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300  
QY 301 SSDLYLVTRHADVIPIVRRRDSRGLSPRPVSYLKGSSGGPLCPGSHAVGIFRAAVCT 360  
Db 301 SSDLYLVTRHADVIPIVRRRDSRGLSPRPVSYLKGSSGGPLCPGSHAVGIFRAAVCT 360  
QY 361 RGVAKAVDFIPVESMETTMR 380  
Db 361 RGVAKAVDFIPVESMETTMR 380

## RESULT 2

US-10-650-585-12  
; Sequence 12, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-650-585-12

Query Match 100.0%; Score 1987; DB 16; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1,2e-193;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIF 60  
Db 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIF 60  
QY 61 DITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLVRKAGHYQMAFMKLAALTG 120  
Db 61 DITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLVRKAGHYQMAFMKLAALTG 120  
QY 121 TYVVDHLTPLODMAHAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
Db 121 TYVVDHLTPLODMAHAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
QY 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKKQVGEVQVSTAT 240  
Db 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKKQVGEVQVSTAT 240  
QY 241 QSFPLATCVNGCVMTVFHAGSKTLAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300  
Db 241 QSFPLATCVNGCVMTVFHAGSKTLAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300  
QY 301 SSDLYLVTRHADVIPIVRRRDSRGLSPRPVSYLKGSSGGPLCPGSHAVGIFRAAVCT 360  
Db 301 SSDLYLVTRHADVIPIVRRRDSRGLSPRPVSYLKGSSGGPLCPGSHAVGIFRAAVCT 360  
QY 361 RGVAKAVDFIPVESMETTMR 380  
Db 361 RGVAKAVDFIPVESMETTMR 380

## RESULT 3

US-10-017-736-11  
; Sequence 11, Application US/10017736

; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 393  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-11

Query Match 100.0%; Score 1987; DB 13; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1,3e-193;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIF 60  
Db 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIF 73  
QY 61 DITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLVRKAGHYQMAFMKLAALTG 120  
Db 74 DITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLVRKAGHYQMAFMKLAALTG 133  
QY 121 TYVVDHLTPLODMAHAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180  
Db 134 TYVVDHLTPLODMAHAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193  
QY 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKKQVGEVQVSTAT 240  
Db 194 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKKQVGEVQVSTAT 253  
QY 241 QSFPLATCVNGCVMTVFHAGSKTLAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300  
Db 254 QSFPLATCVNGCVMTVFHAGSKTLAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 313  
QY 301 SSDLYLVTRHADVIPIVRRRDSRGLSPRPVSYLKGSSGGPLCPGSHAVGIFRAAVCT 360  
Db 314 SSDLYLVTRHADVIPIVRRRDSRGLSPRPVSYLKGSSGGPLCPGSHAVGIFRAAVCT 373  
QY 361 RGVAKAVDFIPVESMETTMR 380  
Db 374 RGVAKAVDFIPVESMETTMR 393

## RESULT 4

US-10-650-585-11  
; Sequence 11, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 393  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-650-585-11

Query Match 100.0%; Score 1987; DB 16; Length 393;

Fri May 7 13:37:02 2004

us-10-650-585-12.rapb

Page 3

Best Local Similarity 100.0%; Pred. No. 1.3e-193;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRAIILLTCAVHELI 60
Db 14 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRAIILLTCAVHELI 73
QY 61 DITLLAIKGLPLVLAQGITKPYFVRAQGLIRACMLVRKAAAGHYVMAFMKLAAL 120
Db 74 DITLLAIKGLPLVLAQGITKPYFVRAQGLIRACMLVRKAAAGHYVMAFMKLAAL 133
QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIISDMVEVKIITWGADTAACGDIISGLPVSAR 180
Db 134 TYVVDHLTPLODMAHAGLRDLAVAVEPVIISDMVEVKIITWGADTAACGDIISGLPVSAR 193
QY 181 GREILLGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
Db 194 GREILLGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 253
QY 241 QSFPLATCNGVCWTVFPGAGSKTLAAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300
Db 254 QSFPLATCNGVCWTVFPGAGSKTLAAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 313
QY 301 SSDIYLVTRADVIYVRRRGRSGSLSPRPVSYLKSGSGPLCPGSHAVGIFRAAVCT 360
Db 314 SSDIYLVTRADVIYVRRRGRSGSLSPRPVSYLKSGSGPLCPGSHAVGIFRAAVCT 373
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 374 RGVAKAVDFIPVESMETTMR 393
```

RESULT 5  
US-10-017-736-2

```
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017, 736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256, 031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2
```

Query Match 100.0%; Score 1987; DB 13; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1.4e-193;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRAIILLTCAVHELI 60
Db 18 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRAIILLTCAVHELI 77
QY 61 DITLLAIKGLPLVLAQGITKPYFVRAQGLIRACMLVRKAAAGHYVMAFMKLAAL 120
Db 78 DITLLAIKGLPLVLAQGITKPYFVRAQGLIRACMLVRKAAAGHYVMAFMKLAAL 137
QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIISDMVEVKIITWGADTAACGDIISGLPVSAR 180
Db 138 TYVVDHLTPLODMAHAGLRDLAVAVEPVIISDMVEVKIITWGADTAACGDIISGLPVSAR 197
QY 181 GREILLGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
Db 198 GREILLGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 257
QY 241 QSFPLATCNGVCWTVFPGAGSKTLAAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300
```

Best Local Similarity 100.0%; Pred. No. 1.4e-193;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 258 QSFPLATCNGVCWTVFPGAGSKTLAAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 317
QY 301 SSDIYLVTRADVIYVRRRGRSGSLSPRPVSYLKSGSGPLCPGSHAVGIFRAAVCT 360
Db 318 SSDIYLVTRADVIYVRRRGRSGSLSPRPVSYLKSGSGPLCPGSHAVGIFRAAVCT 377
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 378 RGVAKAVDFIPVESMETTMR 397
```

RESULT 6  
US-10-650-585-2

```
; Sequence 2, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650, 585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017, 736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256, 031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-2
```

Query Match 100.0%; Score 1987; DB 16; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1.4e-193;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRAIILLTCAVHELI 60
Db 18 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRAIILLTCAVHELI 77
QY 61 DITLLAIKGLPLVLAQGITKPYFVRAQGLIRACMLVRKAAAGHYVMAFMKLAAL 120
Db 78 DITLLAIKGLPLVLAQGITKPYFVRAQGLIRACMLVRKAAAGHYVMAFMKLAAL 137
QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIISDMVEVKIITWGADTAACGDIISGLPVSAR 180
Db 138 TYVVDHLTPLODMAHAGLRDLAVAVEPVIISDMVEVKIITWGADTAACGDIISGLPVSAR 197
QY 181 GREILLGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
Db 198 GREILLGPADNFEQGRRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 257
QY 241 QSFPLATCNGVCWTVFPGAGSKTLAAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 300
Db 258 QSFPLATCNGVCWTVFPGAGSKTLAAGPKGPIITOMYTNVDDLVGMQAPPGARSMTPTCTG 317
QY 301 SSDIYLVTRADVIYVRRRGRSGSLSPRPVSYLKSGSGPLCPGSHAVGIFRAAVCT 360
Db 318 SSDIYLVTRADVIYVRRRGRSGSLSPRPVSYLKSGSGPLCPGSHAVGIFRAAVCT 377
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 378 RGVAKAVDFIPVESMETTMR 397
```

RESULT 7  
US-10-029-907-3

```
; Sequence 3, Application US/10029907
; Publication No. US20020142350A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
```

```

/ TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
/ FILE REFERENCE: 13/083
/ CURRENT APPLICATION NUMBER: US/10/029,907
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,857
/ PRIOR FILING DATE: 2000-12-22
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 2201
/ TYPE: PRT
/ ORGANISM: HCV
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: 882
/ OTHER INFORMATION: Xaa is Lys or Arg
/ NAME/KEY: VARIANT
/ LOCATION: 1489
/ OTHER INFORMATION: Xaa is Leu
US-10-029-907-3

```

```

Query Match          95.0%; Score 1887; DB 13; Length 2201;
Best Local Similarity 93.4%; Pred. No. 2,3e-182;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```

```

QY 2 LITLSPYKVLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFD 61
DB 19 LITLSPHYKFLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFT 78
QY 62 ITKLALIFGPIMLWLOAGITKVPYFRAAGLIRACMLVRAAGHYQMAFMKLAALTGT 121
DB 79 ITKLALIFGPIMLWLOAGITKVPYFRAAGLIRACMLVRAAGHYQMAFMKLAALTGT 138
QY 122 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 181
DB 139 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 198
QY 182 REILGPADNFEQGMRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 241
DB 199 REILGPADNFEQGMRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 258
QY 242 SFLATCVNGVCMVTFEAGSKTLAGPKGPIITQMTNVNDQDLVGMQAPGARSWTPTCTCGS 301
DB 259 SFLATCVNGVCMVTFEAGSKTLAGPKGPIITQMTNVNDQDLVGMQAPGARSWTPTCTCGS 318
QY 302 SDLYLTVRHADVIPIVRRGDSRGSLLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 319 SDLYLTVRHADVIPIVRRGDSRGSLLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 378
QY 362 GVAKAVDFIVPESMETTR 380
DB 379 GVAKAVDFIVPESMETTR 397

```

```

RESULT 8
US-10-309-561-3
/ Sequence 3, Application US/10309561
/ Publication No. US20030148348A1
/ GENERAL INFORMATION:
/ APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
/ TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
/ FILE REFERENCE: 13/083
/ CURRENT APPLICATION NUMBER: US/10/309,561
/ CURRENT FILING DATE: 2002-12-04
/ PRIOR APPLICATION NUMBER: US/10/029,907
/ PRIOR FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,857
/ PRIOR FILING DATE: 2000-12-22
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3

```

```

/ LENGTH: 2201
/ TYPE: PRT
/ ORGANISM: HCV
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: 882
/ OTHER INFORMATION: Xaa is Lys or Arg
/ NAME/KEY: VARIANT
/ LOCATION: 1489
/ OTHER INFORMATION: Xaa is Leu
US-10-309-561-3

```

```

Query Match          95.0%; Score 1887; DB 14; Length 2201;
Best Local Similarity 93.4%; Pred. No. 2,3e-182;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```

```

QY 2 LITLSPYKVLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFD 61
DB 19 LITLSPHYKFLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFT 78
QY 62 ITKLALIFGPIMLWLOAGITKVPYFRAAGLIRACMLVRAAGHYQMAFMKLAALTGT 121
DB 79 ITKLALIFGPIMLWLOAGITKVPYFRAAGLIRACMLVRAAGHYQMAFMKLAALTGT 138
QY 122 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 181
DB 139 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 198
QY 182 REILGPADNFEQGMRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 241
DB 199 REILGPADNFEQGMRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 258
QY 242 SFLATCVNGVCMVTFEAGSKTLAGPKGPIITQMTNVNDQDLVGMQAPGARSWTPTCTCGS 301
DB 259 SFLATCVNGVCMVTFEAGSKTLAGPKGPIITQMTNVNDQDLVGMQAPGARSWTPTCTCGS 318
QY 302 SDLYLTVRHADVIPIVRRGDSRGSLLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 319 SDLYLTVRHADVIPIVRRGDSRGSLLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 378
QY 362 GVAKAVDFIVPESMETTR 380
DB 379 GVAKAVDFIVPESMETTR 397

```

```

RESULT 9
US-10-467-000-1
/ Sequence 1, Application US/10467000
/ Publication No. US20040067486A1
/ GENERAL INFORMATION:
/ APPLICANT: De Francesco, Raffaele
/ APPLICANT: Magliaceto, Giovanni
/ APPLICANT: Pacesa, Giacomo
/ TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
/ FILE REFERENCE: ITR0003P
/ CURRENT APPLICATION NUMBER: US/10/467,000
/ CURRENT FILING DATE: 2003-07-21
/ PRIOR APPLICATION NUMBER: PCT/EP02/00526
/ PRIOR FILING DATE: 2002-01-16
/ PRIOR APPLICATION NUMBER: 60/263,479
/ PRIOR FILING DATE: 2001-01-23
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 3010
/ TYPE: PRT
/ ORGANISM: Con 1 HCV isolate nucleic acid
US-10-467-000-1

```

```

Query Match          95.0%; Score 1887; DB 12; Length 3010;
Best Local Similarity 93.4%; Pred. No. 3,5e-182;

```



Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTTSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGGRDAIILLTCVABEILFD 61  
 DB 828 LTTSPHYKLFARLIMWLYLITRABEHLQWIPPLNVRGGRDAIILLTCABEILFT 887

QY 62 ITKLILAIPLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 121  
 DB 888 ITKLILAIPLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 947

QY 122 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARG 181  
 DB 948 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARG 1007

QY 182 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 241  
 DB 1008 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 1067

QY 242 SFLATCVNGVCMVTFHAGSKTLGAPKGPITQMTYNVDODLVGMQAPPGARSMTPTCTGS 301  
 DB 1068 SFLATCVNGVCMVTFHAGSKTLGAPKGPITQMTYNVDODLVGMQAPPGARSMTPTCTGS 1127

QY 302 SDLYLVTRHADVIIVRRRGDSRGLSPRVSYLKSGSGPILCPSGHAAGIFRAAVCTR 361  
 DB 1128 SDLYLVTRHADVIIVRRRGDSRGLSPRVSYLKSGSGPILCPSGHAAGIFRAAVCTR 1187

QY 362 GVAKAVDFIPVESMETMR 380  
 DB 1188 GVAKAVDFIPVESMETMR 1206

RESULT 10  
 US-09-919-901-4  
 / Sequence 4, Application US/09919901  
 / Publication No. US20030082518A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Potts, Karen E.  
 / APPLICANT: Jackson, Roberta L.  
 / APPLICANT: Patrick, Amy K.  
 / TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 / TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 / FILE REFERENCE: 0125-0005A  
 / CURRENT FILING DATE: 2001-08-02  
 / PRIOR APPLICATION NUMBER: 09/263,933  
 / PRIOR FILING DATE: 1999-02-08  
 / PRIOR APPLICATION NUMBER: 09/129,611  
 / PRIOR FILING DATE: 1998-08-05  
 / NUMBER OF SEQ ID NOS: 33  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 4  
 / LENGTH: 1692  
 / TYPE: PRT  
 / ORGANISM: Artificial Sequence  
 / FEATURE:  
 / OTHER INFORMATION: :  
 / US-09-919-901-4

Query Match 94.7%; Score 1881; DB 10; Length 1692;  
 Best Local Similarity 93.1%; Pred. No. 6,4e-182;  
 Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTTSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGGRDAIILLTCVABEILFD 61  
 DB 107 LTTSPYKVLARLIMWLYLITRABEHLQWIPPLNVRGGRDAIILLTCVABEILFD 166

QY 62 ITKLILAIPLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 121  
 DB 167 ITKLILAIPLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 226

QY 122 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARG 181  
 DB 227 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARG 286

QY 182 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 241  
 DB 287 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 346

QY 242 SFLATCVNGVCMVTFHAGSKTLGAPKGPITQMTYNVDODLVGMQAPPGARSMTPTCTGS 301  
 DB 347 SFLATCVNGVCMVTFHAGSKTLGAPKGPITQMTYNVDODLVGMQAPPGARSMTPTCTGS 406

QY 302 SDLYLVTRHADVIIVRRRGDSRGLSPRVSYLKSGSGPILCPSGHAAGIFRAAVCTR 361  
 DB 407 SDLYLVTRHADVIIVRRRGDSRGLSPRVSYLKSGSGPILCPSGHAAGIFRAAVCTR 466

QY 362 GVAKAVDFIPVESMETMR 380  
 DB 467 GVAKAVDFIPVESMETMR 485

RESULT 11  
 US-10-191-966-4  
 / Sequence 4, Application US/10191966  
 / Publication No. US20030175692A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Potts, Karen E.  
 / APPLICANT: Jackson, Roberta L.  
 / APPLICANT: Patrick, Amy K.  
 / TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 / TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 / FILE REFERENCE: 0125-0005A  
 / CURRENT FILING DATE: 2002-07-10  
 / PRIOR APPLICATION NUMBER: US/09/263,933  
 / PRIOR FILING DATE: 1999-03-08  
 / PRIOR APPLICATION NUMBER: 09/129,611  
 / PRIOR FILING DATE: 1998-08-05  
 / NUMBER OF SEQ ID NOS: 33  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 4  
 / LENGTH: 1692  
 / TYPE: PRT  
 / ORGANISM: Artificial Sequence  
 / FEATURE:  
 / OTHER INFORMATION: :  
 / US-10-191-966-4

Query Match 94.7%; Score 1881; DB 14; Length 1692;  
 Best Local Similarity 93.1%; Pred. No. 6,4e-182;  
 Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTTSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGGRDAIILLTCVABEILFD 61  
 DB 107 LTTSPYKVLARLIMWLYLITRABEHLQWIPPLNVRGGRDAIILLTCVABEILFD 166

QY 62 ITKLILAIPLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 121  
 DB 167 ITKLILAIPLMVLQAGITKVPYFVRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 226

QY 122 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARG 181  
 DB 227 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARG 286

QY 182 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 241  
 DB 287 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLTRDKNOVEGEVQVSTATQ 346

QY 242 SFLATCVNGVCMVTFHAGSKTLGAPKGPITQMTYNVDODLVGMQAPPGARSMTPTCTGS 301  
 DB 347 SFLATCVNGVCMVTFHAGSKTLGAPKGPITQMTYNVDODLVGMQAPPGARSMTPTCTGS 406

QY 302 SDLYLVTRHADVIIVRRRGDSRGLSPRVSYLKSGSGPILCPSGHAAGIFRAAVCTR 361  
 DB 407 SDLYLVTRHADVIIVRRRGDSRGLSPRVSYLKSGSGPILCPSGHAAGIFRAAVCTR 466







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 / Search time 47.9132 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-13

Perfect score: 1842

Sequence: 1 AHQWLPPLNVRGGRDAII.....RGVAKAVDFIPVSMETMR 352

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1842	100.0	352	5	ABG32186 HCV prote
2	1842	100.0	380	5	ABG32185 HCV prote
3	1842	100.0	393	5	ABG32184 HCV prote
4	1842	100.0	409	5	ABG32181 HCV prote
5	1778	96.5	341	5	ABG32187 HCV prote
6	1773	96.3	3010	2	AAR82694 Partial H
7	1768	96.0	3010	2	AAR68622 HCV prote
8	1767	95.9	3010	2	AAR68864 Hepatitis
9	1766	95.9	2201	5	ABG30601 Hepatitis
10	1766	95.9	2201	5	ABG30591 Hepatitis
11	1766	95.9	2201	5	ABG30600 Hepatitis
12	1766	95.9	2201	5	ABG30581 Hepatitis
13	1766	95.9	2201	5	ABG30581 Hepatitis
14	1766	95.9	2201	5	ABG30582 Hepatitis
15	1766	95.9	2201	5	ABG30582 Hepatitis
16	1766	95.9	2201	5	ABG30587 Hepatitis
17	1766	95.9	2201	5	ABG30589 Hepatitis
18	1766	95.9	2201	5	ABG30594 Hepatitis
19	1766	95.9	2201	5	ABG30589 Hepatitis
20	1766	95.9	2201	5	ABG30589 Hepatitis
21	1766	95.9	3010	5	ABG32458 Hepatitis
22	1766	95.9	3010	5	ABG32458 Hepatitis
23	1766	95.9	3010	5	ABG32451 Hepatitis
24	1766	95.9	3010	5	ABG32455 Hepatitis
25	1766	95.9	3010	5	ABG32457 Hepatitis

26	1766	95.9	3010	5	ABG32460 Hepatitis
27	1766	95.9	3010	5	ABG32461 Hepatitis
28	1766	95.9	3010	5	ABG32454 Hepatitis
29	1766	95.9	3011	5	ABG32456 Hepatitis
30	1763	95.7	2201	5	ABG30586 Hepatitis
31	1763	95.7	2201	5	ABG30589 Hepatitis
32	1763	95.7	2201	5	ABG30583 Hepatitis
33	1763	95.7	2201	5	ABG30588 Hepatitis
34	1762	95.7	2201	5	ABG30590 Hepatitis
35	1760	95.5	3010	5	ABG32452 Hepatitis
36	1759	95.5	2201	5	ABG30584 Hepatitis
37	1759	95.5	2201	5	ABG30602 Hepatitis
38	1759	95.5	3010	5	ABG32453 Hepatitis
39	1758	95.4	768	2	AAR40223 Recombina
40	1757	95.4	2307	3	AAY70064 Recombina
41	1755	95.3	2201	5	ABG30585 Hepatitis
42	1754	95.2	2307	3	AAY70065 Recombina
43	1748	94.9	3010	2	AAR98022 Infectiou
44	1748	94.9	3010	4	AAB59174 Protein e
45	1748	94.9	3010	4	AAB31170 Amino aci

## ALIGNMENTS

RESULT 1  
ABG32186 standard; protein; 352 AA.  
ID ABG32186  
XX  
AC ABG32186;  
XX  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX HCV protease NS2/3 truncation mutant 855-1206.  
DE  
XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
XX hepatocellular carcinoma; anti-hepatitis C virus; lauryldithylamine oxide; LDHO;  
XX chaotropic agent; mutant; mutein.  
XX  
XX Hepatitis C virus.  
OS Synthetic.  
OS  
XX MO200248375-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 13-DEC-2001; 2001WO-CA001796.  
XX  
XX 15-DEC-2000; 2000US-0256031P.  
XX  
XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI; 2002-599511/64.  
XX  
XX Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX  
XX Claim 41; Page 61-62; 67pp; English.  
XX  
XX The invention relates to an isolated polypeptide consisting of a full-  
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
XX its truncation or a mutated sequence, where the protease is in a solution  
XX comprising a sufficient concentration of lauryldithylamine oxide (LDHO)  
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABE332198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation mutant 855-1206  
CC (numbered relative to the full length NS2/3 protein)  
CC  
CC  
CC Sequence 352 AA;

```
Query Match      100.0%; Score 1842; DB 5; Length 352;
Best Local Similarity 100.0%; Pred. No. 3.7e-170;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	AHLQWIPPLNVAAGSDAIIILLTCVAHPELIDITKLLALFGPLMVLQAGITKVPEYVR	60
Db	1	AHLQWIPPLNVAAGSDAIIILLTCVAHPELIDITKLLALFGPLMVLQAGITKVPEYVR	60
QY	61	AAGLLIACMLVRRAAAGHYVQMAFPKMLALTSTTYTVEHLTLODMAHAGLBDLVAANEV	120
Db	61	AAGLLIACMLVRRAAAGHYVQMAFPKMLALTSTTYTVEHLTLODMAHAGLBDLVAANEV	120
QY	121	IFSDMEVKIITWGADTPAACGDIISGHPVSARRGRSILLGPADNFEQGWRLAPITASQ	180
Db	121	IFSDMEVKIITWGADTPAACGDIISGHPVSARRGRSILLGPADNFEQGWRLAPITASQ	180
QY	181	QTEGLLGCITTSITGDKNVEGEVQVNSTAQSLATCNNGCWYTFHAGSSTTLAEPK	240
Db	181	QTEGLLGCITTSITGDKNVEGEVQVNSTAQSLATCNNGCWYTFHAGSSTTLAEPK	240
QY	241	GPITQWYTNVDDLVGMQAPFGARSMTPTCGSSDLTYLVTHADVIVERRRDSRGSLLS	300
Db	241	GPITQWYTNVDDLVGMQAPFGARSMTPTCGSSDLTYLVTHADVIVERRRDSRGSLLS	300
QY	301	PRVSTLTKSSSGGPLLCSGHAVGIFFRAVACTRGAAKAVDPIPVESNETMYR	352
Db	301	PRVSTLTKSSSGGPLLCSGHAVGIFFRAVACTRGAAKAVDPIPVESNETMYR	352

XX	ABG32185	standard; protein; 380 AA.
XX	ABG32185;	
XX	05-NOV-2002	(first entry)
XX	HCV protease NS2/3	truncation mutant 827-1206.
XX	HCV;	enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
KW	chronic liver disease;	cirrhosis; end-stage liver disease; vironcide;
KW	hepatotropic; antiinflammatory;	lauryldiethylamine oxide; LDAO;
KW	chaotropic agent; mutant;	mutenin.
XX	Hepatitis C virus.	
OS	Synthetic.	
XX		

PN WO200248375-A2.  
PM  
PP 20-JUN-2002.  
PQ  
PR 13-DEC-2001; 2001WO-CAN001796.  
PS  
PT 15-DEC-2000; 2000US-0256031P.  
PZ  
QA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
QB  
QC Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
QD WPI; 2002-559511/64.  
QE  
QF Novel polypeptide for screening inhibitors of non-structural proteases  
QG useful as therapeutic agents against hepatitis C virus, comprises full  
QH length non-structural protease, or its truncation.  
QI  
QJ Claim 41; Page 60-61; 67pp; English.  
QK

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-termini a residue amino acid 810 to 906, or having a minimal amino acid sequence from residue 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldimethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as A8G31198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterization of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 827-1206 (numbered relative to the full length NS2/3 protein)

SQ Sequence 380 AA;

Query Match	100.0%;	Score 1842;	DB 5;	Length 380;
Best Local Similarity	100.0%;	Pred. No. 4.2e-170;		
Matches 352; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY 1AHQWPIPIPLNNGGRDAIILLTCVAHEHLLFDITKLLAIIFGLMVLQIGITVPEYVR 60

Db 29 AHHQWPIPIPLNNGGRDAIILLTCVAHEHLLFDITKLLAIFGLMVLQIGITVPEYVR 88

QY 61 AAGLLRACMLVYKKAAGHYVQMAFEFKLAALGTIVYDHLTPLODMAHAGLRDLAVANPEV 120

Db 83 AAGLLRACMLVYKKAAGHYVQMAFEFKLAALGTIVYDHLTPLODMAHAGLRDLAVANPEV 148

QY 121 IFSDMEVKIIITMGADTAAAGDIIISGLPVSARGRBIIILGPADNFEQGMRLAPITAYSQ 180

Db 149 IFSDMEVKIIITMGADTAAAGDIIISGLPVSARGRBIIILGPADNFEQGMRLAPITAYSQ 208

QY 181 QTRGLIGCIIISLTERDANKVEBAYQVUSTATQSFLATCVNGVCWTFHQAGSKTLAGPK 240

Db 209 QTRGLGCTITSLTGRDKNQVEGEVSTATQSFLLATCVNGVCMTEFHGSGSKTLAEPK 268  
 Qy 241 GPITQMTNVDDVDMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRSGSGLLS 300  
 Db 269 GPITQMTNVDDVDMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRSGSGLLS 328  
 Qy 301 PRPVSYLKGSSGGLPLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 352  
 Db 329 PRPVSYLKGSSGGLPLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 380

RESULT 3  
 ABG32184  
 ID ABG32184 standard; protein, 393 AA.  
 XX  
 AC ABG32184;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 truncation mutant 815-1206.  
 XX  
 KM HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN W0200248375-A2.  
 XX  
 PD 20-JUN-2002.  
 XX  
 PF 13-DEC-2001; 2001MO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS  
 PS Claim 41; Page 59-60; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-  
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 residue amino acid 810 to 906, or having a minimal amino acid sequence  
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 NS2/3 protease. Also included are (1) a composition (C) comprising an  
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 its truncation or a mutated sequence, where the protease is in a solution  
 comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 appearing as ABG32184; (3) producing (M1) a reformed, inactive HCV NS2/3  
 protease, involving isolating the isolated protease in the presence of a  
 chaotropic agent, refolding the isolated protease by contacting it with a  
 reducing agent, and LDAO in the presence of reduced concentration of the  
 chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 protease, involving diluting refolded inactive NS2/3 protease in a medium  
 containing an activation detergent to induce auto-cleavage of NS2/3  
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 protease, involving incubating the active NS2/3 protease produced by M2  
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 cleavage products or their fragments, and measuring the presence or  
 absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an

CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein).  
 CC  
 XX Sequence 393 AA;  
 XX

Query Match 100.0%; Score 1842; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-170;  
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHLQVWIPPLNVNRGGRDAIILITCAVHPELIPDITKLLIAFGPLMVLQAGITKYPYFVR 60  
 Db 42 AHLQVWIPPLNVNRGGRDAIILITCAVHPELIPDITKLLIAFGPLMVLQAGITKYPYFVR 101  
 Qy 61 AAGLIRACMLVNRKAGGHVQMAFMKLAALTGTYYDHTPLQDMAHAGLRPLAVAVEPV 120  
 Db 102 AAGLIRACMLVNRKAGGHVQMAFMKLAALTGTYYDHTPLQDMAHAGLRPLAVAVEPV 161  
 Qy 121 IFSDMEVKIITWGAADTAACGDIISGLPVARSRGREILGPDNPFQGGWRLAPITAYSQ 180  
 Db 162 IFSDMEVKIITWGAADTAACGDIISGLPVARSRGREILGPDNPFQGGWRLAPITAYSQ 221  
 Qy 181 QTRGLGCTITSLTGRDKNQVEGEVSTATQSFLLATCVNGVCMTEFHGSGSKTLAEPK 240  
 Db 222 QTRGLGCTITSLTGRDKNQVEGEVSTATQSFLLATCVNGVCMTEFHGSGSKTLAEPK 281  
 Qy 241 GPITQMTNVDDVDMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRSGSGLLS 300  
 Db 282 GPITQMTNVDDVDMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRSGSGLLS 341  
 Qy 301 PRPVSYLKGSSGGLPLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 352  
 Db 342 PRPVSYLKGSSGGLPLCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 393

RESULT 4  
 ABG32181  
 ID ABG32181 standard; protein, 409 AA.  
 XX  
 AC ABG32181;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 (810-1206).  
 XX  
 KM HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 398..409  
 FT /note="Streptavidin tag"  
 XX  
 PN W0200248375-A2.  
 XX  
 PD 20-JUN-2002.  
 XX  
 PF 13-DEC-2001; 2001MO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 PI MPI: 2002-599511/64.  
 XX N-PSDB; ABRK90406.  
 XX

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 useful as therapeutic agents against hepatitis C virus, comprises full  
 length non-structural protease, or its truncation.  
 PT

PS Claim 42; Fig 1B; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
 XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
 XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 XX its truncation or a mutated sequence, where the protease is in a solution  
 XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 XX appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3  
 XX protease, involving isolating the isolated protease by contacting it with a  
 XX chaotropic agent, and LDAO in the presence of reduced concentration of the  
 XX chaotropic agent, and LDAO in the presence of reduced concentration of the  
 XX chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 XX protease, involving diluting refolded inactive NS2/3 protease in a medium  
 XX containing an activation detergent to induce auto-cleavage of the NS2/3  
 XX protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 XX protease, involving incubating the active NS2/3 protease and produce  
 XX for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 XX cleavage products or their fragments, and measuring the presence or  
 XX absence of uncleaved NS2/3 protease, and measuring the presence or  
 XX ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 XX active NS2/3 protease, involving carrying out M3 in the presence of, or  
 XX absence of the potential inhibitor, comparing the amount of uncleaved  
 XX NS2/3 protease, cleavage products or their fragments. The protease is  
 XX useful for detailed biochemical characterisation of the enzymes and in  
 XX the development of in vitro assays for screening novel inhibitors of  
 XX NS2/3 protease which are useful as therapeutic agents against HCV  
 XX infection (which causes chronic liver disease, cirrhosis and end-stage  
 XX liver disease. M1 is useful for high level production of protease. The  
 XX present sequence represents the NS2/3 (810-1206) protein, which has a C-  
 XX terminal streptavidin tag

Sequence 409 AA:

Query Match 100.0%; Score 1842; DB 5; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-170;  
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AHQVWIPPLNVRGRDAIILLTCVAPHEIFDITKLLAIFGFLMVLQGITKVPFVR 60  
 46 AHQVWIPPLNVRGRDAIILLTCVAPHEIFDITKLLAIFGFLMVLQGITKVPFVR 105  
 61 AGLIRACMLVRKAGHYVOMAFMKLAALTGTVVVHPLPLDMWARGRLDAVVEPV 120  
 106 AGLIRACMLVRKAGHYVOMAFMKLAALTGTVVVHPLPLDMWARGRLDAVVEPV 155  
 121 IFSDMEVKITWGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITVYSQ 180  
 166 IFSDMEVKITWGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITVYSQ 225  
 181 QTRGLIGCITSLTGRDNQVEGEVYSTATQSFPLATCNGVWYTFPHAGSSTTLAEPK 240  
 226 QTRGLIGCITSLTGRDNQVEGEVYSTATQSFPLATCNGVWYTFPHAGSSTTLAEPK 285  
 241 GPITOMYNNVODLVGMQAPPGARSMPTCTGSSDLYLTFRHADYIVFRERGDSRGSLLS 300  
 286 GPITOMYNNVODLVGMQAPPGARSMPTCTGSSDLYLTFRHADYIVFRERGDSRGSLLS 345  
 301 PRPVSYLKSSGGPLLCPSGHAAGVIFRAAVCTRGVAAVDFIPVESMETTMR 352

DB 346 PRPVSYLKSSGGPLLCPSGHAAGVIFRAAVCTRGVAAVDFIPVESMETTMR 397

RESULT 5  
 ABG32187  
 ID ABG32187 standard; protein; 341 AA.

AC ABG32187;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 866-1206.

XX HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 XX hepatocellular carcinoma; anti-inflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.

OS Hepatitis C virus.  
 OS Synthetic.

PN WO200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

PI MPI: 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 useful as therapeutic agents against hepatitis C virus, comprises full  
 length non-structural protease, or its truncation.  
 PT

PS Claim 41; Page 62-63; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
 XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
 XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 XX its truncation or a mutated sequence, where the protease is in a solution  
 XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 XX appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3  
 XX protease, involving isolating the isolated protease by contacting it with a  
 XX chaotropic agent, and LDAO in the presence of reduced concentration of the  
 XX chaotropic agent, and LDAO in the presence of reduced concentration of the  
 XX chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 XX protease, involving diluting refolded inactive NS2/3 protease in a medium  
 XX containing an activation detergent to induce auto-cleavage of the NS2/3  
 XX protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 XX protease, involving incubating the active NS2/3 protease and produce  
 XX for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 XX cleavage products or their fragments, and measuring the presence or  
 XX absence of uncleaved NS2/3 protease, and measuring the presence or  
 XX ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 XX active NS2/3 protease, involving carrying out M3 in the presence of, or  
 XX absence of the potential inhibitor, comparing the amount of uncleaved  
 XX NS2/3 protease, cleavage products or their fragments. The protease is  
 XX useful for detailed biochemical characterisation of the enzymes and in  
 XX the development of in vitro assays for screening novel inhibitors of  
 XX NS2/3 protease which are useful as therapeutic agents against HCV  
 XX infection (which causes chronic liver disease, cirrhosis and end-stage  
 XX liver disease. M1 is useful for high level production of protease. The



CC present sequence represents the NS2/3 truncation mutant 866-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 341 AA;

Query Match 96.5%; Score 1778; DB 5; Length 341;

Best Local Similarity 100.0%; Pred. No. 5,9e-164; Mismatches 0; Indels 0; Gaps 0;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 VRGSRDAIILITCAVHPELIFDITKLLAFGLMTQAGITKYPYFVRAQGLIRACMLV 71  
 Db 1 VRGSRDAIILITCAVHPELIFDITKLLAFGLMTQAGITKYPYFVRAQGLIRACMLV 60  
 QY 72 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDLAVAVEPVIFSDMEVKIIT 131  
 Db 61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDLAVAVEPVIFSDMEVKIIT 120  
 QY 132 WGAADTAACGDIISGLPVASARRGRIILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 191  
 Db 121 WGAADTAACGDIISGLPVASARRGRIILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180  
 QY 192 SLTRDKNOVEGEVQVSTATQSFATCVNGVCTVTFHAGSKTLAGPKPITQMTNV 251  
 Db 181 SLTRDKNOVEGEVQVSTATQSFATCVNGVCTVTFHAGSKTLAGPKPITQMTNV 240  
 QY 252 QDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVIPVRRGDSRGLSLFRPVSYLKSS 311  
 Db 241 QDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVIPVRRGDSRGLSLFRPVSYLKSS 300  
 QY 312 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 352  
 Db 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 341

#### RESULT 6

AAR82694 ID AAR82694 standard; protein; 3010 AA.

XX AAR82694;

AC 16-OCT-2003 (revised)  
 DT 14-NOV-1996 (first entry)

XX Partial HCV non-structural polyprotein.

XX proteinase; hepatitis C virus; screening; inhibitor; proteolytic;  
 KM identification; cleavage.

XX Hepatitis C virus; Virus.

XX Key Location/Qualifiers

FT Protein 898..1233  
 FT /note="partial proteinase; see AAR82692"

FT Protein 992..1907  
 FT /note="partial proteinase; see AAR82693"

XX JF07184648-A.

XX 25-JUL-1995.

XX 05-FEB-1993; 93JP-00018854.

XX 07-FEB-1992; 92JP-00022657.

XX 18-SEP-1992; 92JP-00249240.

XX 04-DEC-1992; 92JP-00325303.

XX (KAEN/) KAENNO K.

XX (SUMO) SUMITOMO METAL IND LTD.

XX (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

XX WPI; 1995-287962/38.

XX N-PSDB; AAT03960.

PT An HCV proteinase active substance - which has activity as an anti-HCV  
 PT agent and can be used to screen for proteinase inhibitors.

XX Disclosure; Page 39-48; 52pp; Japanese.

CC The present sequence is a partial Hepatitis C Virus (HCV) polyprotein  
 CC from the non-structural region. Partial proteinase sequences (AAR82692-  
 CC 93) are contained within this sequence. The proteinases can be used as  
 CC anti-HCV agents. They can also be used to screen cpds. for their ability  
 CC to inhibit their proteolytic activity. In this way proteinase inhibitors  
 CC can be identified. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 3010 AA;

Query Match 96.3%; Score 1773; DB 2; Length 3010;

Best Local Similarity 94.6%; Pred. No. 4e-162; Mismatches 7; Indels 0; Gaps 0;

Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHQVWIPPLNRRGGRDAIILITCAVHPELIFDITKLLAFGLMTQAGITKYPYFV 60  
 Db 855 AHQVWIPPLNRRGGRDAIILITCAVHPELIFDITKLLAFGLMTQAGITKYPYFV 914  
 QY 61 AAGLRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDLAVAVEPV 120  
 Db 915 AAGLRACMLVRAKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDLAVAVEPV 974  
 QY 121 IFSDEVEKIITWGAADTAACGDIISGLPVASARRGRIILGPADNFEQGMRLAPITAYSQ 180  
 Db 975 IFSDEVEKIITWGAADTAACGDIISGLPVASARRGRIILGPADNFEQGMRLAPITAYSQ 1034  
 QY 181 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCVNGVCTVTFHAGSKTLAGPK 240  
 Db 1035 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCVNGVCTVTFHAGSKTLAGPK 1094  
 QY 241 GPITQMTNVDDQDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVIPVRRGDSRGLSL 300  
 Db 1095 GPITQMTNVDDQDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVIPVRRGDSRGLSL 1154  
 QY 301 PRPVSYLKSSGSGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 352  
 Db 1155 PRPVSYLKSSGSGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 1206

#### RESULT 7

AAR68622 ID AAR68622 standard; protein; 3010 AA.

XX AAR68622;

AC 16-OCT-2003 (revised)  
 DT 16-OCT-1995 (first entry)

XX HCV protein cleavable with new serine proteinase.

XX proteinase; serine; cleavage; hepatitis C virus; HCV.

XX Hepatitis C virus; Virus.

XX Key Location/Qualifiers

FT Cleavage-site 2419..2420  
 FT /note="Serine protease cleavage site"

XX JF06315377-A.

XX 15-NOV-1994.

XX 06-MAY-1993; 93JP-00105666.

XX 06-MAY-1993; 93JP-00105666.

XX (KAEN/) KAENNO K.

XX (SUMO) SUMITOMO METAL IND LTD.

XX (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

XX WPI; 1995-032330/05.  
 DR N-PSDB; AAQ80498.  
 XX  
 PT New HCV-originated proteinase active substance - used for site-specific  
 cleavage by an intermolecular reaction and the purification thereof.  
 XX  
 PS Disclosure; Page 10-19; 23pp; Japanese.  
 XX  
 CC This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is  
 CC cleaved between amino acids 2419 and 2420, by a new serine protease,  
 CC contg. the sequence of AAR68621. The proteinase is purified as a fused  
 CC product with the dihydrofolate reductase protein by using a methotrexate  
 CC column. It can be used for the development of an inhibitor for HCV  
 CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)  
 CC  
 XX  
 SQ Sequence 3010 AA;  
 Query Match 96.0%; Score 1769; DB 2; Length 3010;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;  
 Matches 332; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 AHLQWVLPPLNVNRRGRDAIILLTCVHPELIFDITKLLAFGLPMLVLAQGITRVYFVR 60  
 DB 855 AHLQWVLPPLNVNRRGRDAIILLTCVHPELIFDITKLLAFGLPMLVLAQGITRVYFVR 914  
 QY 61 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAAVAVESV 120  
 DB 915 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAAVAVESV 974  
 QY 121 IFSMEVKIITWGDITACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180  
 DB 975 VFSDMETKLIITWGDITACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 1034  
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFIATCNGVCMVTFHAGSKTLAGPK 240  
 DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFIATCNGVCMVTFHAGSKTLAGPK 1094  
 QY 241 GPITQWNTNVDQDLYGMAPEPARGASMTPTCTGSSSDIYLVTRADVIPVRRRGDSRGLIS 300  
 DB 1095 GPITQWNTNVDQDLYGMAPEPARGASMTPTCTGSSSDIYLVTRADVIPVRRRGDSRGLIS 1154  
 QY 301 PRPISTYKGGSSGGPLCPSGHANGIFPRAVCTRGVAKAVFIVESEMTTMR 352  
 DB 1155 PRPISTYKGGSSGGPLCPSGHANGIFPRAVCTRGVAKAVFIVESEMTTMR 1206  
 RESULT 8  
 AAR68864 standard; protein: 3010 AA.  
 ID AAR68864;  
 XX  
 AC AAR68864;  
 XX  
 DT 06-DEC-1995 (first entry)  
 XX  
 DE Hepatitis C virus RNA helicase.  
 XX  
 XX Hepatitis C virus, HCV; non-A non-B; helicase gene; RNA helicase;  
 KM baculovirus; recombinant production.  
 XX  
 OS Hepatitis C virus.  
 XX  
 XX  
 FT Key  
 FT Location/Qualifiers  
 FT 196..198  
 FT /label= N-linked glycosylation site  
 FT 209..211  
 FT /label= N-linked glycosylation site  
 FT 234..236  
 FT /label= N-linked glycosylation site  
 FT 250..252  
 FT /label= N-linked glycosylation site  
 FT 303..307  
 FT /label= N-linked glycosylation site

FT Region 325..327  
 FT /label= N-linked glycosylation site  
 FT Region 417..419  
 FT /label= N-linked glycosylation site  
 FT Region 423..425  
 FT /label= N-linked glycosylation site  
 FT Region 430..432  
 FT /label= N-linked glycosylation site  
 FT Region 448..450  
 FT /label= N-linked glycosylation site  
 FT Region 532..534  
 FT /label= N-linked glycosylation site  
 FT Region 556..558  
 FT /label= N-linked glycosylation site  
 FT Region 576..578  
 FT /label= N-linked glycosylation site  
 FT Region 623..625  
 FT /label= N-linked glycosylation site  
 FT Region 645..647  
 FT /label= N-linked glycosylation site  
 FT Region 1213..1215  
 FT /label= N-linked glycosylation site  
 FT Region 1255..1257  
 FT /label= N-linked glycosylation site  
 FT Region 2041..2043  
 FT /label= N-linked glycosylation site  
 FT Region 2077..2079  
 FT /label= N-linked glycosylation site  
 FT Region 2240..2242  
 FT /label= N-linked glycosylation site  
 FT Region 2788..2790  
 FT /label= N-linked glycosylation site  
 FT Region  
 PN JP06319583-A.  
 XX  
 PD 22-NOV-1994.  
 XX  
 PR 18-SEP-1992; 92JP-00249241.  
 XX  
 PR 18-SEP-1992; 92JP-00249241.  
 XX  
 PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  
 XX  
 DR WPI; 1995-040330/06.  
 DR N-PSDB; AAQ81559.  
 XX  
 PT of hepatitis C virus helicase gene in baculovirus - useful for large  
 PT scale prodn. of RNA helicase.  
 XX  
 PS Claim 1; Fig 1-4; 9pp; Japanese.  
 XX  
 CC AAQ81559 encodes AAR6864 hepatitis C virus (HCV) RNA helicase. The DNA  
 CC was used in the construction of an expression vector, which was used to  
 CC transform a baculovirus host. The transformed baculovirus could then be  
 CC used for the recombinant prodn. of HCV RNA helicase  
 XX  
 SQ Sequence 3010 AA;  
 Query Match 95.9%; Score 1767; DB 2; Length 3010;  
 Best Local Similarity 94.3%; Pred. No. 1.5e-161;  
 Matches 332; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 AHLQWVLPPLNVNRRGRDAIILLTCVHPELIFDITKLLAFGLPMLVLAQGITRVYFVR 60  
 DB 855 AHLQWVLPPLNVNRRGRDAIILLTCVHPELIFDITKLLAFGLPMLVLAQGITRVYFVR 914  
 QY 61 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAAVAVESV 120  
 DB 915 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAAVAVESV 974  
 QY 121 IFSMEVKIITWGDITACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180  
 DB 975 VFSDMETKLIITWGDITACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 1034

QY 181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAEPK 240  
 Db 1035 QTRGLGCIITSLTGRDKNQVDEGEVQLSTATQSFATCVNGVCMVTFHAGSKTLAEPK 1094  
 QY 241 GPIQMTNTVDODLVGMQAPPGARSMPTCTGSSDLYLTRHADVTPVRRRDSRSLLS 300  
 Db 1095 APIQMTNTVDODLVGMQAPPGARSMPTCTGSSDLYLTRHADVTPVRRRDSRSLLS 1154  
 QY 301 PRPVSYLKGSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETTR 352  
 Db 1155 PRPVSYLKGSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETTR 1206  
 RESULT 9  
 ABG30601  
 ID ABG30601 standard; protein; 2201 AA.  
 AC ABG30601;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT Misc-difference 2183 /label= Arg, Lys  
 FT Misc-difference 2183 /note= "Wild type Met substituted by Thr"  
 FT  
 XX WO200252015-A2.  
 XX 04-JUL-2002.  
 XX 20-DEC-2001; 2001MO-CA001843.  
 XX 22-DEC-2000; 2000US-0257857P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PI Kukolj G, Pause A;  
 XX WPI; 2002-575382/61.  
 DR  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G12042/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon AEGK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in

CC the claims of the invention  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 AHQWIPPLNVRGGRDAIILITCAVHPELLFDITKLALIFGPMVLQAGITKVPYVR 60  
 Db 46 AHQWIPPLNVRGGRDAIILITCAVHPELLFDITKLALIFGPMVLQAGITKVPYVR 105  
 QY 61 AAGLIRACMLVAKAAGHVVQAFMKLAALITGYVDHLPLODAHAGRLAVAVEPV 120  
 Db 106 AAGLIRACMLVAKAAGHVVQAFMKLAALITGYVDHLPLODAHAGRLAVAVEPV 165  
 QY 121 IFSDMEVKIITWGAOTTAACGDIISGLPVASARGREIILGPADNFEQGQWRLAPITAVSQ 180  
 Db 166 VPSDMETKVIITWGAOTTAACGDIISGLPVASARGREIILGPADNFEQGQWRLAPITAVSQ 225  
 QY 181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAEPK 240  
 Db 226 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAEPK 285  
 QY 241 GPIQMTNTVDODLVGMQAPPGARSMPTCTGSSDLYLTRHADVTPVRRRDSRSLLS 300  
 Db 286 GPIQMTNTVDODLVGMQAPPGARSMPTCTGSSDLYLTRHADVTPVRRRDSRSLLS 345  
 QY 301 PRPVSYLKGSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETTR 352  
 Db 346 PRPVSYLKGSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETTR 397

RESULT 10  
 ABG30591  
 ID ABG30591 standard; protein; 2201 AA.  
 AC ABG30591;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 751  
 FT Misc-difference 882 /note= "Wild type Ser substituted by Gly"  
 FT Misc-difference 882 /label= Arg, Lys  
 FT  
 XX WO200252015-A2.  
 XX 04-JUL-2002.  
 XX 20-DEC-2001; 2001MO-CA001843.  
 XX 22-DEC-2000; 2000US-0257857P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PI Kukolj G, Pause A;  
 XX WPI; 2002-575382/61.  
 DR  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX

Claim 3; Page; 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NT<sub>5'</sub>), where guanine at position 1 is substituted for adenine, a HCV polypeptide region coding for a HCV polypeptide, and a 3'-UTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NT<sub>5'</sub>, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon ApeX2 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention

Sequence 2201 AA;

every Match 95.9%; Score 1766; DB 5; Length 2201;

```

      Local Similarity  94.3%;  Pred. No. 1.2e-16;
      Mismatches      11;      Indels      0; Gaps      0
      Conservative     9;

```

atches	332;	Conservative	11;	Mismatches	9;	Indels	0;	Gaps	0.
--------	------	--------------	-----	------------	----	--------	----	------	----

1 AHLQWAEPLLVNVRGGRDAIILTLCAVHPELIFITITLALIGPMVLQAGITKVPYFVR 60  
46 AHLQWAEPLLVNVRGGRDAVILTLCAIHPELIFITITLALIGPMVLQAGITKVPYFVR 105  
61 AGLIRACMLVRKAGSHYVQVAFMKLALITGVYVDHLTQIDKBAHAGLADLVANAEV 120  
106 AHGIRACMLVRKAGSHYVQVAMLMKLAALITGVYVDHLTQIDKBAHAGLADLVANAEV 165  
121 IFSDMEVKITIWGADTACCDIISGLPVSARSGREILGPDNFEQOQWRLLAPITAYSQ 180  
166 VPSDMETKVIITWGADTLACGDIILGLPVSARSGREILGPDNLSGQOQWRLLAPITAYSQ 225  
181 QTRGLIGCITISLGRDKNOVEGEVYVYVATNOSPLATGVNVCMTVHAGSGKTLGPK 240  
226 QTRGLIGCITISLGRBRNDVEGEVYVYVATNOSPLATGVNVCMTVHAGSGKTLGPK 285  
241 GBITOMYTNVQDILVQOAPPGARSMPTCTGSSDLYLTVRHADYIPVRRGDSRGSLLS 300  
286 GBITOMYTNVQDILVQOAPPGARSLTPTCGSSDLYLTVRHADYIPVRRGDSRGSLLS 345  
301 PEPVSYLKGSSGGPLTCSGHAIVGIPFAAACITGVAKADLPVPSMETMTR 352  
346 PEPVSYLKGSSGGPLTCSGHAIVGIPFAAACITGVAKADLPVPSMETMTR 397

LT 11  
0600  
ABG30600 standard; protein; 2201 AA

ABG30600;

21-OCT-2002 (first entry)

Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

Self-replicating; hepatitis C virus; HCV, HCV replication inhibitor  
cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutuin.

Hepatitis C virus.  
Synthetic.

Key	Location/Qualifiers

Misc-difference 882

11

```
/note= "Wild type Pro substituted by Leu"
```

XX W0200252015-A2  
PN

PD 04-JUL-2002

PF 20-DEC-2001; 2001WO-CA001843

PR 22-DEC-2000; 2000US-0257857P

PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD

PI Kukolj G, 'Pause A;

DR WPI; 2002-575382/61

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.

PS Claim 3; Page: 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV), a polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G12042/C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence and/or represents a mutant of the hepatitis C virus replicon AP6C12 and contains the viral protease NS3/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in Ab630580 using information given in the claims of the invention

Sequence 2201 AA;

Query Match	95.9%	Score 1766	DB 5	Length 2201
-------------	-------	------------	------	-------------

Best Local Similarity 94.3%; Pred. No. 1.2e-161;  
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0

1 AHLQWIPRLNRCGRDAIILLTCVAHPELFDITKLLAIFGPIMLQAGITKVYFVR 60

```

Db      46  AHUQWVIRPUNVGRDAVILITCAIHBEFLTITKLLAILGELVWLQAGIKTFVYR  105
QY      61  AAGLIRACMLVRRKAAGHYVQMAFKLTAALTGYVYDHLTPLODMAHAGIRDLAVVEV  120
Db      106 AHGIRACMLVRRKVGSHYVQWALMKLAALTGYVYDHLTPLRMAHAGIRDLAVVEV  165
QY      121 IFSDEWVKIITMGADTAAACGDIISGLPYRSARGREILGPSADNEGSGWMLAPITAYEQ  180
Db      166 VFSDEWETKIITMGADTAAACGDIISGLPYRSARGREILGPSADNEGSGWMLAPITAYEQ  225
QY      181 QTEGLLGCIITSLTGRKXQVEGEVQVVSATQSPFATCNVGCWYTFHAGSKTLAGRK  240
Db      226 QTGGLLGCIITSLTGRDRQVEGEVQVVSATQSPFATCNVGCWYTFHAGSKTLAGRP  285
QY      241 GPLTQWYTNVDOLVGMQAPGARSMPTCTGSSDILVTRHADVLPVRRGRSGRGLLS  300
Db      286 GPLTQWYTNVDOLVGMQAPGARSLTCTGSSDILVTRHADVLPVRRGRSGRGLLS  345
QY      301 PRVSYLKSGSGGPLLCPSGAHVIFPRAVCTRGVAAKVDFIPVESHETIMR  352
Db      346 PRVSYLKSGSGGPLLCPSGAHVIFPRAVCTRGVAAKVDFIPVESHETIMR  397

```

RESULT 1  
ABG30581



Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60  
 DB 46 AHLQWIPPLNVGRDAVILLTCALHPELIFITTKLLAILGLMVLQAGITKVPYFVR 105  
 QY 61 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVYDHLTPLODMHAGRLDAVAPEV 120  
 DB 106 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVYDHLTPLODMHAGRLDAVAPEV 165  
 QY 121 IFSDMEVKIITMGADTAAAGDIIISGLPVARSRGREIILGPADNFEQGWLLAIPITAYSQ 180  
 DB 166 VPSDMETKVTITMGADTAAAGDIIISGLPVARSRGREIILGPADNFEQGWLLAIPITAYSQ 225  
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFALTQVNCVCTVYFHAGSKTLTAGPK 240  
 DB 226 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFALTQVNCVCTVYFHAGSKTLTAGPK 285  
 QY 241 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIYVRRRGDSRGLLS 300  
 DB 286 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIYVRRRGDSRGLLS 345  
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFPRAVCTRGVAKAVDFIPVSMETTM 352  
 DB 346 PRPVSYLKSSGGPILCPSGHAGVIFPRAVCTRGVAKAVDFIPVSMETTM 397

RESULT 14  
 ABG30582  
 ID ABG30582 standard; protein; 2201 AA.  
 XX ABG30582;  
 AC  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /note= "Wild type Lys substituted by Lys or Arg"  
 FT Misc-difference 1233  
 FT /note= "Wild type Gly substituted by Cys"  
 FT  
 XX  
 PN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 PI WPI: 2002-575382/61.  
 DR N-PSDB; ABR8574.  
 XX  
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Disclosure; Page 59-69; 140pp: English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-

CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apgk12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon Apgk12 shown in ABG30581

Sequence 2201 AA:

Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60  
 DB 46 AHLQWIPPLNVGRDAVILLTCALHPELIFITTKLLAILGLMVLQAGITKVPYFVR 105  
 QY 61 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVYDHLTPLODMHAGRLDAVAPEV 120  
 DB 106 AAGLIRACMLVRKAAAGHYVQMAFMKLAALTGYVYDHLTPLODMHAGRLDAVAPEV 165  
 QY 121 IFSDMEVKIITMGADTAAAGDIIISGLPVARSRGREIILGPADNFEQGWLLAIPITAYSQ 180  
 DB 166 VPSDMETKVTITMGADTAAAGDIIISGLPVARSRGREIILGPADNFEQGWLLAIPITAYSQ 225  
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFALTQVNCVCTVYFHAGSKTLTAGPK 240  
 DB 226 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFALTQVNCVCTVYFHAGSKTLTAGPK 285  
 QY 241 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIYVRRRGDSRGLLS 300  
 DB 286 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIYVRRRGDSRGLLS 345  
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFPRAVCTRGVAKAVDFIPVSMETTM 352  
 DB 346 PRPVSYLKSSGGPILCPSGHAGVIFPRAVCTRGVAKAVDFIPVSMETTM 397

RESULT 15  
 ABG30580  
 ID ABG30580 standard; protein; 2201 AA.  
 XX ABG30580;  
 AC  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /note= "Encoded by ARG"  
 FT  
 XX  
 PN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843;  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX

PI Kukolj G, Pause A;  
yy

DR WPI; 2002-575382/61.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
PT possess enhanced transduction or replication efficiency, useful for  
PT evaluating potential inhibitors of HCV replication.

PS Disclosure; Page 69-74; 140pp; English

CC The invitation describes a self-replicating hepatitis C virus (HCV).  
CC The polynucleotide molecule comprising a 5'-non translated region (5'NT),  
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
CC useful for efficiently establishing cell culture replication. The self-  
CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
CC position 1 is substituted for A, and therefore provides an alternative to  
CC existing systems comprising a self-replicating HCV RNA molecule that, in  
CC conjunction with mutations in the HCV non-structural region, such as the  
CC G(2042)/C/R mutations, transduces and/or replicates with greater  
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
CC region NS2/3 and contains the viral protease NS2/3, protease complex  
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

Sequence 2201 AA;

Query Match 95.9%; Score 1766; DB 5; Length 2201;

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY	46	AHQQWIPRLNPNVGGSDAIIILTCANPHELIDITIKLLAFGLPMVQAQITVPEVFR	60
Db	46	AHQQWIPRLNPNVGGSDAIIILTCANPHELITIKLLAFGLPMVQAQITVPEVFR	1050
QY	61	AQGLIPACMLVRKAAGHYVQAFMKLALITGVYDHLTPIQWANAAGLIDLAFAVEPV	120
Db	106	AHQLIPACMLVRKAAGHYVQAMKMLALITGVYDHLTPIRLWANAAGLIDLAFAVEPV	165
QY	121	IFSDMEKTIITWAGDPAACGGDIISGLPVARAREGREILIGPANDNEGOSWRLAPITYSQ	180
Db	166	VFSDMETKVIITWAGDPAACGGDIILGLPVARAREGREILIGPANDSLEGOMRLAPITYSQ	225
QY	181	QITGGLIGCIITSLITGEDKNQVBGEVQVNSTAQSLATLCVNGVCMVYFHGAGSKTLA	240
Db	226	QITGGLIGCIITSLITGEDKNQVBGEVQVNSTAQSLATLCVNGVCMVYFHGAGSKTLA	285
QY	241	GPITQWYTNVDQDLVWQAPAPARASMTPTCGSSDLVYVTRHADVIPIRRRGDSRGSL	300
Db	286	GPITQWYTNVDQDLVWQAPAPARASLTPTCGSSDLVYVTRHADVIPIRRRGDSRGSL	345
QY	301	PRRVSITLKSSGGPRLICPSGHAIVGIFRAAVCTRGYAKAVDPIPVESMETTMR	352
Db	346	PRRVSITLKSSGGPRLICPSGHAIVGIFRAAVCTRGYAKAVDPIPVESMETTMR	397

Search completed: May 6, 2004, 09:30:47  
Job time : 48.9132 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 11.4639 Seconds

(Without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-13

Sequence: 1 AHQVWIPPLNVRGRDAII.....RGVAKAVDFIPVESMETTVR 352

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 293366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 293366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR:78:\*  
2: PIR:1:\*  
3: PIR:3:\*  
4: PIR:4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1773	96.3	3010	1 GNMVUJ	genome polypeptide
2	1766	95.9	3010	1 A45573	genome polypeptide
3	1749	95.0	3010	1 GNMVUW	genome polypeptide
4	1717	93.2	3010	1 S18030	genome polypeptide
5	1699	92.2	3010	1 GNMVUJ	genome polypeptide
6	1616	87.7	3011	1 GNMVU3	genome polypeptide
7	1610	87.4	3011	1 S40770	genome polypeptide
8	1605	87.1	3011	1 GNMVUJ	genome polypeptide
9	1401	76.1	3014	1 JCS430	genome polypeptide
10	1303	70.7	3033	1 GNMVU8	genome polypeptide
11	1301	70.6	3033	1 UQ1303	genome polypeptide
12	408.5	22.2	3005	1 T08841	polypeptide - dour
13	342.5	18.6	2970	2 T08839	polypeptide - marm
14	102.5	5.6	660	2 VHMWH2	structural protein
15	102.5	5.6	692	2 H71426	hypothetical prote
16	101	5.5	564	2 S36637	signal recognition
17	101	5.5	600	2 E46642	DNA-directed DNA p
18	100.5	5.5	353	2 G87392	conserved hypochet
19	99	5.4	399	2 AH3038	conserved hypochet
20	99	5.4	1085	2 C98247	hypothetical 50.8k
21	97.5	5.3	1085	2 T03531	coBN protein homol
22	95.5	5.2	470	2 TC4098	tetracycline 6-hyd
23	94.5	5.1	2796	2 TC4743	fatty-acid synthas
24	93.5	5.1	1380	2 T18309	receptor-adenylate
25	93	5.0	7463	2 T36248	CDA peptide synthet
26	92.5	5.0	706	2 S33761	transferrin precur
27	92.5	5.0	716	2 G83612	hypothetical prote
28	92	5.0	659	2 B44212	structural protein
29	91.5	5.0	3414	1 GNMVUJ	genome polypeptide

30	91	4.9	504	2 A84212	hypothetical prote
31	90.5	4.9	868	2 H81775	aconitate hydratase
32	90.5	4.9	871	2 H72597	hypothetical prote
33	90.5	4.9	3069	2 H70656	fatty-acid synthase
34	89.5	4.9	961	2 AE0375	probable cation-tr
35	89	4.8	446	2 AP1509	conserved hypochet
36	88.5	4.8	652	2 C70688	probable nitrate r
37	88.5	4.8	3076	2 A87058	fatty acid synthase
38	87.5	4.8	3412	1 GNMVUJ	genome polypeptide
39	86.5	4.7	347	2 S44167	malate dehydrogena
40	86.5	4.7	451	2 H82044	C4-dicarboxylate t
41	86.5	4.7	470	1 NM1VW8	exo-alpha-sialidas
42	86.5	4.7	707	2 D84154	cadmium-transport
43	86	4.7	338	1 S34984	UDPglucose 4-epime
44	85.5	4.6	348	2 H70549	probable pdhb prot
45	85	4.6	470	1 NM1V9	exo-alpha-sialidas

## ALIGNMENTS

## RESULT 1

GNMVUJ genome polypeptide - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

Query Match 96.3%; Score 1773; DB 1; Length 3010;

Best Local Similarity 94.6%; Pred. No. 6,8e-14;

Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

QY	1 AHQVWIPPLNVRGRDAIIITLCVHPELFDIKLAIIFGPMVQAGTKYVPR 60
DB	855 AHQVWIPPLNVRGRDAIIITLCVHPELFDIKLAIIFGPMVQAGTKYVPR 914
QY	61 AAGLIRACLVKRAAGGHVQVAFMKALALGTYYVDHITPLQDVAHAGLRDLAAVPEV 120
DB	915 AAGLIRACLVKRAAGGHVQVAFMKALALGTYYVDHITPLQDVAHAGLRDLAAVPEV 974

QY	121	FSDMEVYIIMWADPLAACGDIISGLPVASARGRIILLGPADNFGGOCWRLAITAYSQ	180
D6	975	VFSDMETKILITMGADPLAACGDIISGLPVASARGKEILLGPADSGEGOCWRLAITAYSQ	1034
QY	181	QTRGLGCIILSLTGRDNQVEGEVQVSTATOSPLACVNGCMTVPHGAGSKTLGPK	240
D6	1035	QTRGLGCIILSLTGRDNQVDEGVQVSTATQSPFLACVNGCMTVYHAGSKTLGPK	1094
QY	241	GPITDMYTNVQDILVGMQAPPGARSMTPCTGSSSLYLVTTHADYIPRRRGDSRGSILS	300
D6	1095	GPITDMYTNVQDILVGMQAPPGARSMTPCTGSSSLYLVTTHADVVPRRRGDSRGSILS	1154
QY	301	PRPVSYLKGSSGGPILCPSGHAVGIFRAAVCTRGAKAVDFIIPESMETTKR	352
D6	1155	PRPISYLVKGSSGGPILCPSGHAVGIFRAAVCTRGAKAVDFIIPESMETTKR	1206

## RESULT 2

genome polyprotein - hepatitis C virus (strain J7)  
N.Contains: capsid protein C, envelope protein M, hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a), nonstructural protein NS4b, nonstructural protein NS5  
C.Species: hepatitis C virus  
C.Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C.Accession: A45573  
R.Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.T.; Nakazawa, T.; Hijikata, Y. *Virus Res.* 23, 39-53, 1992  
A.Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  
A.Reference number: A45573, MUID:92295714, PMID:1318627  
A.Accession: A45573  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-3010 <TAN>  
A.Cross-references: GB:D11168; GB:D01171; NID:g2212612; PIDN:BA01943.1; PID:g2212613  
A.Experimental source: HCV-JT  
A.Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:P.106207)  
C.Superfamily: hepatitis C virus genome polyprotein  
C.Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
P.2-115/Product: capsid protein C #status predicted <CPC>  
F.115-191/Product: envelope protein M #status predicted <EM>  
F.192-388/Product: major envelope protein E #status predicted <ME>  
F.390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F.730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F.1007-1615/Product: hepacivirin #status predicted <NS3>  
F.1200-1237/Region: nucleotide-binding motif A (P-loop)  
F.1312-1317/Region: nucleotide-binding motif B  
F.1316-1319/Region: DXH motif  
F.1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F.1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F.2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match	95.9%	Score 1766;	DB 1;	Length 3010;
Best Local Similarity	94.9%	Pred. No. 2.7e-140;		
Matches 334;	Conservative 7;	Mismatches 11;	Indels 0;	Gaps 0;
QY	1	AHLQWIPPLNVRGRDAIILLTCAMPELIFDITKLLAIFGRLPMUQGITKVPFVR	60	
Db	855	AHLQWVPEPLNVRGRDAIILLTCAMPELIFDITKLLAIFGRLPMUQAATIMPFYVR	914	
QY	61	AQGLIRACMLVRKAAGHYVQMAFMKLAALIGTYVYDHLTLEJODMAHAGRLDAVAEVP	120	
Db	915	AQGLIRACMLVRKAYAGHYVQMAFMKLAALIGTYVYDHLTLEJODMAHAGRLDAVAEVP	974	
QY	121	IFSDMEVKIITWGDITAAQDITISGLPIVSARGREIILGPDNFEQGGWRLAIPITAYSQ	180	
Db	975	VFSDMEETKIITWGDITAAQDITIGLPIVSARGREIILGPDNFEQGGWRLAIPITAYAQ	1034	
QY	181	QTRGLGCIITSLTRDKNQYEGEVQVSTQSFSLATCVAGVQCVTFVHGAGSKTLAPK	240	
Db	1035	QTRGLGCIITVSLTRDKNQYEGEVQVSTQSFSLATCVAGVQCVTFVHGAGSKTLAPK	1094	
QY	241	GPITMTYNNVDQDLVGQAPPGARSMTEPCTGSSDLIVYTHADIVPIRRKGBDSRGSILS	300	

Dd	1095	GPIQMTNVQDLVGMHAPFGAASLIPCTGSSDYLVTFRHAAVLPFRRRDGGSLLS	115
Qy	301	PRPVSYLKSGSGGPLCPGSHAVGI FRAAVCTRGAAVDEI PVESMETTR	352
Dd	1155	PRPVSYLKSGSGGPLCTSGHAAVGI FRAAVCTRGAAVNDI PVESMETTR	1206

### RESULT 3

genome polyprotein - hepatitis C virus (strain Taiwan)  
 NC\_001845: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C.Species: hepatitis C virus  
 ANote: host Homo sapiens (man)  
 C.Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #ext\_c-change 19-Jan-2001  
 C.Accession: A40244  
 R.Chen, P.J., Lin, M.H., Tai, K.F., Liu, P.C., Lin, C.J., Chen, D.S.  
 Virology 188, 102-113, 1992  
 A>Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A.Reference number: A40244 / MIMD:92230206; EMD:114449

A.Molecule type: genomic RNA  
A.Residues: 1-3010 <CHE>  
A.CrossRef: references: GB:M84754  
C.Superfamily: hepatitis C virus genome polypeptide  
C.Keywords: Afp; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F.1-115/Product: capsid protein C #status predicted <CP>  
F.116-151/Product: envelope protein M #status predicted <EPM>  
F.192-369/Product: major envelope protein E #status predicted <ME>  
F.730-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F.730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F.1007-1615/Product: hepatitisvirin #status predicted <NS3>  
F.1120-1237/Region: nucleotide-binding motif A (P-loop)  
F.1132-1317/Region: nucleotide-binding motif B  
F.1136-1319/Region: DEXH motif  
F.1186-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
F.1863-2013/Product: nonstructural protein NS4 #status predicted <NS4>  
F.1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F.196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match	95.0%	Score 1749	DB 1	Length 3010
Best Local Similarity	92.9%	Pred. No. 7, 2e-139		
Matches 327/	Conservative 14	Mismatches 11	Indels 0	Gaps 0

  

QY	1	AHLQWVLPPLNVNRGSDAIIILLTCANHPSELIPDITLLKLAIGPLMLQAGSTTKPYPR	60
DB		.....:.....	
QY	855	AHLQWVLPPLNVNRGSRDAIILLTCAAPSELIPDITKLLAIGPLMLVQAGSTRIPYPR	914
DB		.....:.....	
QY	61	AQGLIRACMLVRRKAAAGHYQMAFMKLAALTGYVVVDHLTPLQDMVHAGLRDLAAVEV	120
DB	915	AQGLIRACMLVRRKAAAGHYQMALMKLAALTGYVVVDHLTPLQDMVHAGLRDLAAVEV	974
QY	121	IFSMERKIIITWGAADPLAACGDIISGLPVSGAPRREITLGPANNEGCGRLLAPLTAYSQ	180
DB	975	VFSMEKRIITWGAADPLAACGDIISGLPVSGARRREITLGPANSLSEGRMLLAPLTAYAQ	1034
QY	181	QTRGLLGCIIITSLTGDARDKQVGESEVQVVSSTATQSFATCVNGVCMTFHGAGSKTLAGPK	240
DB	1035	QTRGLPFCIIITSLTGDARDKQVGESEVQVVSSTATQSFATCINGCVMTHGAGSKTLAGPK	1094
QY	241	GPITQMTNVNDQDLVGMQAPPGARSMTPCCGSSDLYLTLRADYIVPVRRGDSGSLDS	300
DB	1095	GPITQMTNVNDQDLVGMHAPQGRSLTLPCTCGSSDLYLTLRADYIVPVRRGDSGSLDS	1154
QY	301	PRPVSYIKGSSGGPILCPSGHAVGIFRAAVCTRGVAKANDPVPESMETMTR	352
DB	1155	PRPISYLKSSGGPILCPSGHVVGIFRAAVCTRGVAKAVDFVFSMETMTR	1206

  

RESULT 4	518030	genome	polyprotein - hepatitis C virus (isolate JX1)
----------	--------	--------	---

N/Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
A/Variety: isolate JX1  
C/Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 23-Mar-2001  
C/Accession: S18030; S33570; A88332; S18029  
R/Honda, M.; Kaneko, S.; Maseh, U.; Kodayashi, K.; Murakami, S.  
Submitted to the EMBL Data Library, September 1991  
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated from a single patie  
A/Reference number: S18028  
A/Accession: S18030  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <HON>  
A/Cross-references: EMBL:X61596; NID:959478; PIDN:CAA43793.1; PID:959479  
A/Experimental source: isolate JX1 from an individual  
R/Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
A/Reference number: A48332; MUID:91119270; PMID:8380322  
A/Accession: S33570  
A/Molecule type: genomic RNA  
A/Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
A/Cross-references: EMBL:X61591  
A/Note: this sequence is inconsistent with the nucleotide translation  
A/Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320  
as Trp, and TTC for residue 771 as Ser  
A/Note: sequence extracted from NCBI database (NCBI:121747, NCBI:121748)  
C/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEP>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif A (P-loop)  
F/1616-1862/Product: DEXH motif  
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS

Query Match 93.2%; Score 1717; DB 1; Length 3010;  
Best Local Similarity 92.3%; Pred. No. 3, 6e-136;  
Matches 325; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

1 AHIQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLOAGITKVPYR 60  
Db 855 ACQWVWPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLOAGITKVPYR 914

QY 61 AOGILPACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAIVEPV 120  
Db 915 AOGILPACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAIVEPV 974

QY 121 IFSDMEVKIITMGADTAACGDIISGLPVSARSGREILGPADNFEQGRLLAPITAYSO 180  
Db 975 VFSDEMKIITMGADTAACGDIISGLPVSARSGREILGPADNFEQGRLLAPITAYSO 1034

QY 181 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAPK 240  
Db 1035 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAPK 1094

QY 241 GPTTQWTVNVDQDLVGMQAPGASMTPTCTGSSDLYLVTHADVIPIRRGDSGSLIS 300  
Db 1095 GPTTQWTVNVDQDLVGMQAPGASMTPTCTGSSDLYLVTHADVIPIRRGDSGSLIS 1154

QY 301 PRPVSYLKGSSGGLPLCPSGHANVITRAVCTRGAKAVDPIVPESMETMR 352  
Db 1155 PRPVSYLKGSSGGLPLCPSGHANVITRAVCTRGAKAVDPIVPESMETMR 1206

GNMWTG  
genome polyprotein - hepatitis C virus  
N/Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text\_change 19-Jan-2001  
C/Accession: A38465  
R/Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J  
J. Virol. 65, 1105-1113, 1991  
A/Title: Structure and organization of the hepatitis C virus genome isolated from human c  
A/Reference number: A38465; MUID:91140698; PMID:1847440  
A/Accession: A38465  
A/Molecule type: genomic RNA  
A/Residues: 1-3010 <YAK>  
A/Cross-references: EMBL:M58335; NID:9329770; PIDN:AAA72945.1; PID:9329771  
A/Superfamily: hepatitis C virus genome polyprotein  
C/Keywords: ATP; capsid protein C #status predicted <CPC>  
F/2-115/Product: capsid protein C #status predicted <CPC>  
F/116-191/Product: envelope protein M #status predicted <EMP>  
F/192-389/Product: major envelope protein E #status predicted <MEP>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif A (P-loop)  
F/1312-1319/Region: DEXH motif  
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 92.2%; Score 1699; DB 1; Length 3010;  
Best Local Similarity 91.5%; Pred. No. 1, 2e-134;  
Matches 322; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

1 AHIQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLOAGITKVPYR 60  
Db 855 ADLHWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLOAGITKVPYR 914

QY 61 AOGILPACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAIVEPV 120  
Db 915 AOGILPACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAIVEPV 974

QY 121 IFSDMEVKIITMGADTAACGDIISGLPVSARSGREILGPADNFEQGRLLAPITAYSO 180  
Db 975 VFSDEMKIITMGADTAACGDIISGLPVSARSGREILGPADNFEQGRLLAPITAYSO 1034

QY 181 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAPK 240  
Db 1035 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCTVFEHAGSKTLAPK 1094

QY 241 GPTTQWTVNVDQDLVGMQAPGASMTPTCTGSSDLYLVTHADVIPIRRGDSGSLIS 300  
Db 1095 GPTTQWTVNVDQDLVGMQAPGASMTPTCTGSSDLYLVTHADVIPIRRGDSGSLIS 1154

QY 301 PRPVSYLKGSSGGLPLCPSGHANVITRAVCTRGAKAVDPIVPESMETMR 352  
Db 1155 PRPVSYLKGSSGGLPLCPSGHANVITRAVCTRGAKAVDPIVPESMETMR 1206

## RESULT 6

GNMWTG  
genome polyprotein - hepatitis C virus (strain HCV-1)  
N/Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 19-Jan-2001  
C/Accession: A39166; PQ0403; PQ0404  
R/Choo, Q.-U.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit  
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A/Title: Genetic organization and diversity of the hepatitis C virus.  
A/Reference number: A39166; MUID:91172826; PMID:1848704  
A/Accession: A39166

A/Molecule type: mRNA  
 A/Residues: 1-3011 <CHO>  
 A/Cross-references: GB:M62321; NID:9329873; PIDN:AAA5676.1; PID:9329874  
 R/Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to H  
 A/Reference number: PQ0393; MUID:92268871; PMID:1136939  
 A/Accession: PQ0403  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CHA>  
 A/Cross-references: DDBJ:D10128  
 A/Experimental source: isolates E-b16  
 A/Accession: PQ0404  
 A/Status: Preliminary  
 A/Molecule type: genomic RNA  
 A/Residues: 1577-1633 <CH2>  
 A/Experimental source: isolates E-b17  
 C/Superfamily: hepatitis C virus genome polypeptide; glycoprotein; hydrolase; nonstructura  
 C/Keywords: ATP; capsid protein; envelope protein; #status predicted <CPC>  
 F:115-191/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 87.7%; Score 1616; DB 1; Length 3011;  
 Best Local Similarity 84.1%; Pred. No. 12e-127;  
 Matches 296; Conservative 28; Mismatches 28; Indels 0; Gaps 0;

QY 1 AHIQWIPPLNVRGGRDAIILLTCVAHPELFDITKLLAIFGPMWLOAGITKVPYFVR 60  
 DB 855 AQLHWVPEPLNVRGGRDAVILLMCVHPTLVFDITKLLAVGPELWLOAGITKVPYFVR 914  
 QY 61 AGLIRACMVRKXAGGHVQYMAFMKLAALTGTYYVDHPLDPMANAGRLDAVAEVP 120  
 DB 915 VQGLRICALARKKVGHHVQYMAITKLGALTGYVNHPLPLDMANAGRLDAVAEVP 974  
 QY 121 IFSDEMYKIIITWGAADTAACGDIISGLPVSARRGREILLGPADNFBEGQRLAPITAYQ 180  
 DB 975 VFSQMETKLITWGAADTAACGDIISGLPVSARRGREILLGPADNFBEGQRLAPITAYQ 1034  
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 240  
 DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 1094  
 QY 241 GPITQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGSLLS 300  
 DB 1095 GPVQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGSLLS 1154  
 QY 301 PRPVSYLKSGSGGPELCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETMR 352  
 DB 1155 PRPISYKSGSGGPELCPAGHAGVIFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 7  
 S40770  
 genome polypeptide - hepatitis C virus  
 N/Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 protein NS4); nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C/Accession: S40770; P01285  
 R/Okamoto, H.  
 submitted to the EMBL Data Library, March 1992  
 A/Reference number: S40770

A/Accession: S40770  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3011 <OKA>  
 A/Cross-references: EMBL:D10749; NID:9221586; PIDN:BA01582.1; PID:9221587  
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,  
 Umi. J. Exp. Med. 60, 167-177, 1990  
 A/Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A/Reference number: P01284; MUID:91013116; PMID:2170712  
 A/Accession: P01285  
 A/Molecule type: genomic RNA  
 A/Residues: 1-513 <OK2>  
 A/Cross-references: GB:P00831; NID:9221511; PIDN:BA00705.1; PID:9221512  
 A/Experimental source: isolate HC-J1  
 C/Superfamily: hepatitis C virus genome polypeptide  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 87.4%; Score 1610; DB 1; Length 3011;  
 Best Local Similarity 83.5%; Pred. No. 4e-127;  
 Matches 294; Conservative 26; Mismatches 30; Indels 0; Gaps 0;

QY 1 AHIQWIPPLNVRGGRDAIILLTCVAHPELFDITKLLAIFGPMWLOAGITKVPYFVR 60  
 DB 855 AQLHWVPEPLNVRGGRDAVILLMCVHPTLVFDITKLLAVGPELWLOAGITKVPYFVR 914  
 QY 61 AGLIRACMVRKXAGGHVQYMAFMKLAALTGTYYVDHPLDPMANAGRLDAVAEVP 120  
 DB 915 VQGLRICALARKKVGHHVQYMAITKLGALTGYVNHPLPLDMANAGRLDAVAEVP 974  
 QY 121 IFSDEMYKIIITWGAADTAACGDIISGLPVSARRGREILLGPADNFBEGQRLAPITAYQ 180  
 DB 975 VFSQMETKLITWGAADTAACGDIISGLPVSARRGREILLGPADNFBEGQRLAPITAYQ 1034  
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 240  
 DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 1094  
 QY 241 GPITQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGSLLS 300  
 DB 1095 GPVQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGSLLS 1154  
 QY 301 PRPVSYLKSGSGGPELCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETMR 352  
 DB 1155 PRPISYKSGSGGPELCPAGHAGVIFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 8  
 G40770  
 genome polypeptide - hepatitis C virus (strain H)  
 N/Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 protein NS4); nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 A/Note: host Homo sapiens (man)  
 C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C/Accession: A36814; A41546  
 R/Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 submitted to GenBank, July 1992  
 A/Description: Genomic structure of the human prototype strain H of hepatitis C virus: CC  
 A/Reference number: A36814  
 A/Accession: A36814  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3011 <INC>

A/Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
 R:Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A/Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
 A/Reference number: A41566; MUID:92052256; PMID:1658800  
 A/Contents: annotation  
 A/Note: neither amino acid nor nucleotide sequence is given  
 A/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:1.15/Product: capsid protein C #status predicted <GPC>  
 F:1.16-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1616/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F:1316-1319/Region: DEXH motif  
 F:1316-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 87.1%; Score 1605; DB 1; Length 3011;  
 Best Local Similarity 83.5%; Pred. No. 1e-126; Indels 0; Gaps 0;  
 Matches 294; Conservative 30; Mismatches 28; Indels 0; Gaps 0;

QY 1 AHHQVWIPPLNVRGGDAIILLTCVAPBELIFDITKLLAIFGFLMWLAGITKVPYFVR 60  
 DB 855 AQLHWVWVPLNVRGGDAVIMLTCLFHPALGPEITKILFGILGPIYLQHSITKVPYFDR 914  
 QY 61 AAGLIRACMLVRKAGHYQVMAFMKLAALTGYVDHLPLQDMAHAGRLDAVAVEPV 120  
 DB 915 VQGLRICALARKLAGHYQVMAFMKLAALTGYVDHLPLQDMAHAGRLDAVAVEPV 974  
 QY 121 IFSDEVKIITWGADTAACGDIISGLPVSARSGREILLGPADNPEGQWRLAPITAYSQ 180  
 DB 975 VFSMEKLTITWGADTAACGDIISGLPVSARSGREILLGPADNPEGQWRLAPITAYSQ 1034  
 QY 181 QTRGLGCIITITSLTRDKNOVEGEVQVSTATOSTLATCNGVCWTFVHAGSKTLAEPK 240  
 DB 1035 QTRGLGCIITITSLTRDKNOVEGEVQVSTATOSTLATCNGVCWTFVHAGSKTLAEPK 1094  
 QY 241 GPITQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRRADVLPVRRGDSRGLLS 300  
 DB 1095 GPVITQTVNDODLVGMOPPGARSMPTCGSSDLVYTRRADVLPVRRGDSRGLLS 1154  
 QY 301 PRPVSYLKSGSGGFLCPGSHAVGIFRAAVCTRGVAVADVFIVESMETTMR 352  
 DB 1155 PRPISYLKSGSGGFLCPGSHAVGIFRAAVCTRGVAVADVFIVENLETTMR 1206

RESULT 9  
 JC5620 genome polyprotein - hepatitis C virus (isolate EUH1480)  
 N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C/Accession: J05620  
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A/Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
 A/Reference number: J05620; MUID:97365593; PMID:9223423  
 A/Accession: J05620  
 A/Molecule type: mRNA  
 A/Residues: 1-3014 <CHA>  
 A/Cross-references: GB:J113184  
 A/Experimental source: genotype 5a, which predominates in South Africa  
 A/Note: the translation of the nucleotide sequence is not complete in this paper  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:384-408/Region: hypervariable #status predicted <MB>  
 F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1008-1616/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
 F:1313-1318/Region: nucleotide-binding motif B  
 F:1317-1320/Region: DEXH motif  
 F:1517-1563/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 76.1%; Score 1401; DB 1; Length 3014;  
 Best Local Similarity 70.7%; Pred. No. 1.8e-109; Indels 0; Gaps 0;  
 Matches 249; Conservative 51; Mismatches 52; Indels 0; Gaps 0;

QY 1 AHHQVWIPPLNVRGGDAIILLTCVAPBELIFDITKLLAIFGFLMWLAGITKVPYFVR 60  
 DB 856 AQLHWVWVPLNVRGGDAVIMLTCLFHPALGPEITKILFGILGPIYLQHSITKVPYFDR 915  
 QY 61 AAGLIRACMLVRKAGHYQVMAFMKLAALTGYVDHLPLQDMAHAGRLDAVAVEPV 120  
 DB 916 ARLRLCLALALVYKGVQALHLGLTGYVDHLPLQDMAHAGRLDAVAVEPV 975  
 QY 121 IFSDEVKIITWGADTAACGDIISGLPVSARSGREILLGPADNPEGQWRLAPITAYSQ 180  
 DB 976 VFSMEKLTITWGADTAACGDIISGLPVSARSGREILLGPADNPEGQWRLAPITAYSQ 1035  
 QY 181 QTRGLGCIITITSLTRDKNOVEGEVQVSTATOSTLATCNGVCWTFVHAGSKTLAEPK 240  
 DB 1036 QTRGVGAIVSLTRDKNOVEGEVQVSTATOSTLATCNGVCWTFVHAGSKTLAEPK 1095  
 QY 241 GPITQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRRADVLPVRRGDSRGLLS 300  
 DB 1096 GPVQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRRADVLPVRRGDSRGLLS 1155  
 QY 301 PRPVSYLKSGSGGFLCPGSHAVGIFRAAVCTRGVAVADVFIVESMETTMR 352  
 DB 1156 PRPISYLKSGSGGFLCPGSHAVGIFRAAVCTRGVAVADVFIVENLETTMR 1207

RESULT 10  
 GNRVJ8 genome polyprotein - hepatitis C virus (strain HC-98)  
 N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C/Species: hepatitis C virus  
 C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C/Accession: A40250; E00397; E00559  
 R:Okamoto, H.; Kurai, K.; Okada, S.T.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;  
 Virology 189, 331-341, 1992  
 A/Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo  
 A/Reference number: A40250; MUID:92230323; PMID:1314459  
 A/Accession: A40250  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3033 <OKR>  
 A/Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609  
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to ex  
 A/Reference number: P00393; MUID:92268871; PMID:1316939  
 A/Accession: P00393  
 A/Molecule type: genomic RNA  
 A/Residues: 2678-2754 <CHA>  
 A/Cross-references: DBJ:D10134  
 A/Experimental source: isolate E-b12  
 R:Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,  
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
 A/Title: Distribution of plural HCV types in Japan.  
 A/Reference number: P00554; MUID:92068504; PMID:1720309  
 A/Accession: P00554

A:Molecule type: tRNA  
 A:Residues: 2678-2729 <KAT>  
 A:Cross-references: GB:ID0562; GB:ID0518; NID:g221523; PIDN:BA001418.1; PID:g221524  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein C #status predicted <EPC>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1011-1619/Product: hepatitis virus #status predicted <NS3>  
 F:1234-1241/Region: nucleotide-binding motif A (P-loop)  
 F:1316-1321/Region: nucleotide-binding motif B  
 F:1320-1323/Region: DEXH motif  
 F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>  
 F:1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,233,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 70.7%; Score 1303; DB 1; Length 3033;  
 Best Local Similarity 65.6%; Pred. No. 3.3e-101;  
 Matches 231; Conservative 56; Mismatches 65; Indels 0; Gaps 0;

QY 1 AHQWIPPLNVRGGRDAILITCAVHPELIPDTIKLLAFGLWLVQAGITKVPFVR 60  
 DB 859 AQLQWVPLVGRGGRGILWAVIILPRLVFEVITMLAIIIGPAVLLKASILRLPVEVR 918  
 QY 61 AAGIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAGRLDAVAEVP 120  
 DB 919 AHALLRCTVYKHLAAGRYVQMLITIGRTGYIIDHLSPLSTWMAAGRLDAVAEVP 978  
 QY 121 IFSDEYKLTITWADTAACGDIISGLPVSARRGREIILGPDNFEQGMRLAPITAYQ 180  
 DB 979 VFSPMEKVIWGAETACGDIILHGLPVSARLRGEVILGPADGYTSKGMKLAPITAYQ 1038  
 QY 181 QTRGLGCIITSLTGRPKNVEGEVQVSTATOSFLATCMVGVTFHAGSKTLAGPK 240  
 DB 1039 QTRGLGCIITSLTGRPKNVEGEVQVSTATOSFLATCMVGVTFHAGSKTLAGPK 1098  
 QY 241 GPITQWTVNDODIVGQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRSGSLIS 300  
 DB 1099 GPITQWTVNDODIVGQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRSGSLIS 1158  
 QY 301 PRPVSYLKSSGGPILCPSGHANGIFPAANCTGVAKAVDFIVESMETMR 352  
 DB 1159 PRPLSTLKSSGGPILCPSGHANGIFPAANCTGVAKAVDFIVESMETMR 1210

RESULT 11  
 QY1303 genome polypeptide - hepatitis C virus (isolate HC-J6)  
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (BC 3.4.21.98) (nonstructu  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C:Accession: J01303  
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.  
 U: Gen. Virol. 72, 2697-2704, 1991  
 A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum  
 A:Reference number: J01303; MUID:92044440; PMID:1658196  
 A:Accession: J01303  
 A:Molecule type: genomic RNA  
 A:Residues: 1-303 <OKA>  
 A:Cross-references: GB:D00944; NID:g221650; PIDN:BA00792.1; PID:g221651  
 A:Experimental source: isolate HC-J6 from a Japanese individual  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polypeptide; serine proteinase; trans  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein C #status predicted <EPC>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1011-1619/Product: hepatitis virus #status predicted <NS3>

F:1316-1321/Region: nucleotide-binding motif B  
 F:1320-1323/Region: DEXH motif  
 F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>  
 F:1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,281

Query Match 70.6%; Score 1301; DB 1; Length 3033;  
 Best Local Similarity 66.5%; Pred. No. 4.9e-101;  
 Matches 234; Conservative 51; Mismatches 67; Indels 0; Gaps 0;

QY 1 AHQWIPPLNVRGGRDAILITCAVHPELIPDTIKLLAFGLWLVQAGITKVPFVR 60  
 DB 859 AMQEMAPVQVGRGGRGILWAVIILPRLVFEVITMLAIIIGPAVLLKASILRLPVEVR 918  
 QY 61 AAGIRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAGRLDAVAEVP 120  
 DB 919 AHALLRCTVYKHLAAGRYVQMLITIGRTGYIIDHLSPLSTWMAAGRLDAVAEVP 978  
 QY 121 IFSDEYKLTITWADTAACGDIISGLPVSARRGREIILGPDNFEQGMRLAPITAYQ 180  
 DB 979 VFSPMEKVIWGAETACGDIILHGLPVSARLRGEVILGPADGYTSKGMKLAPITAYQ 1038  
 QY 181 QTRGLGCIITSLTGRPKNVEGEVQVSTATOSFLATCMVGVTFHAGSKTLAGPK 240  
 DB 1039 QTRGLGCIITSLTGRPKNVEGEVQVSTATOSFLATCMVGVTFHAGSKTLAGPK 1098  
 QY 241 GPITQWTVNDODIVGQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRSGSLIS 300  
 DB 1099 GPITQWTVNDODIVGQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGRSGSLIS 1158  
 QY 301 PRPVSYLKSSGGPILCPSGHANGIFPAANCTGVAKAVDFIVESMETMR 352  
 DB 1159 PRPLSTLKSSGGPILCPSGHANGIFPAANCTGVAKAVDFIVESMETMR 1210

RESULT 12  
 T08841 polypeptide - douroucouli hepatitis GB virus A  
 C:Species: douroucouli hepatitis GB virus A  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C:Accession: T08841  
 R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.T.; Montes, C.C.; Mushahwar, I.K.  
 U: Gen. Virol. 79, 41-45, 1998  
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A:Reference number: Z16486; MUID:99120816; PMID:9460920  
 A:Accession: T08841  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3005 <ERK>  
 A:Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: polypeptide

Query Match 22.2%; Score 408.5; DB 2; Length 3005;  
 Best Local Similarity 31.3%; Pred. No. 1.1e-25;  
 Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 13 RGGRR-----AIIITCAVHPELIPDTIKLLAFGLWLVQAGITKVPFVFAOGLIRAC 68  
 DB 813 RGGRRWRTAVVAVAGIIFPRREVRCASALR-----LAALDSIDYLETL-ILRAA 864  
 QY 69 MLVRA-----AGHYVQMAFMKLAALTGTYYVDHLTPLQDMAHAGRLDAVAEVP 120  
 DB 865 GPAPARALIDSLTFAGDADLITRAFRRLERGGVLLFCHGQGVSGAAAILDLGVALLRPV 924  
 QY 121 IFSDEYKLTITWADTAACGDIISGLPVSARRGREIILGPDNFEQGMRLAPITAYQ 178  
 DB 925 SVTARDCYIVDAATLTCGGVREGLPVVARGDVAVGVPSRALPFGVPPAPVAV- 983  
 QY 179 SQTGLGCIITSLTGRPKNVEGEVQVSTATOSFLATCMVGVTFHAGSKTLAG 238  
 DB 984 MGRGIGFVSVYKTSWLGDESHESIVVLGSTRSMGTGVNGVMTTFHGSNARTLAG 1043

QY 239 PKGPIITOMYNVDDIDVGMQAPPGARSMPTCTCGSSDLYLTVTRADYIPVRRRDSSTGSL 238  
 Db 1044 PVGVNCRWSPSPDVAVYPLPFGASCCEPCCKGTGWCIRN--DQALCHGRISKLVEL 1101  
 QY 299 LSPRPVSYLKSSGGPLICPSGHAAGIFRAAVCTRGV-----AKAVDFIPVES 346  
 Db 1102 DLPEISDPRSSSGSPILCDGCHVGM--VSVLHRGVKTVGRVYKRWETLPPKDS 1155

## RESULT 13

polyprotein - marmoset hepatitis GB virus A  
 C/Species: marmoset hepatitis GB virus A  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C/Accession: T08839  
 R/Eker: J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A/Reference number: Z16486; MUID:98120818; PMID:9460920  
 A/Accession: T08839  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: genomic RNA  
 A/Residues: 1-2970 <BRK>  
 A/Cross-references: EMBL:AF023424; NID:g2828597; PIDN:MAC40501.1; PID:g2828598  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: polyprotein

Query Match 18.6%; Score 342.5; DB 2; Length 2970;  
 Best Local Similarity 28.9%; Pred. No. 4.2e-20;  
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;

QY 37 LLLAIFP-----LMVLQAGITKYPFVRAQGLIRACMLVKKXAGHYVMATMKA 88  
 Db 816 LVAMFWPFEIAAVCAVFIIGFPPVVDILFVLVSSPNVRLARVDSLVAADKXA 875  
 QY 89 ALGTIVYDHLTLOD--WAAG-----LRDLAVERPVSFSDMEVKIITWGADTA 137  
 Db 876 T--TWLVGELKRNCFVYMAQVTRRTAEQLRQGFLEPAVAPEDCAMVRDARFL 932  
 QY 138 ACGDIIISGLPVSARRGRBELLGPADNFEQGWRL-----LAPITXQSQTGLGCIIT 191  
 Db 933 SCQGSVHGKRVYARRGDEVILIGLVN--WELPFGFVPTAVVVA--HKGSGFEGVKT 987  
 QY 192 SLTRGXNOVEGVQVYSTATQSFATCVGVCMVFHSGSKTLAAGPKPIITOMYTNV 251  
 Db 988 SMTGWDETHVGVVVLGSTRMGTGVGVVYTHGSMARLLAQMGPVNSRWMSAS 1047  
 QY 252 ODLVGMQAPPGARSMPTCTCGSSDLYLTVTRADYIPVRRRDSRGSLS----- 300  
 Db 1048 DVAIVYPLVGAACLEPCCKQPOGVVVI-----RND--GALCHGILGRVETDL 1094  
 QY 301 PRPVSYLKSSGGPLICPSGHAAGIFRAAVCTRG-----VAKAVDFIPVESMETT 350  
 Db 1095 PAELCDPRSSSGSPILCDGCHVGM--VSVLHRGSVGTGRTKPKWETLPRALITHT 1150

## RESULT 14

VHMMH2  
 structural protein 2 precursor - hepatitis E virus (strain Burma)  
 C/Species: hepatitis E virus  
 C/Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1999  
 C/Accession: C40778  
 R/Tam: A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.  
 Virology 185, 120-131, 1991  
 A/Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi  
 A/Reference number: A40778; MUID:92024067; PMID:1926770  
 A/Accession: C40778  
 A/Molecule type: genomic RNA  
 A/Residues: 1-660 <TAM>  
 A/Cross-references: GB:M73218; NID:g330023; PIDN:AAA45736.1; PID:g330026  
 A/Note: the authors translated the codon GGC for residue 2 as Ala  
 C/Superfamily: hepatitis E virus structural protein 2

C/Keywords: structural protein  
 F/1-22/Domain: signal sequence #status predicted <SIG>  
 F/23-660/Product: structural protein 2 #status predicted <SP2>

Query Match 5.6%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 1.2;  
 Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

QY 35 TWLL--AIFGPMVLQAG-----ITKPYFVR--AQGLIRACMLVKKXAGHYVMA 83  
 Db 151 TWLVYAAPLSPILPQDGTNTHIMATSNVYAQVRVAPATIRYRPLVNNAGVAIGIS 210  
 QY 84 FKKALATGTYYDHTPLQDMAHGLDIAVAVPVFSDMEVKIITWGADTAACGDI 143  
 Db 211 FWPQTTTFTSV-----DNMSITSDVRLVQPGIASBELVI----- 246  
 QY 144 SGLPVSARRGRBELLGPAD--NFEQGWRLAPI--TAVSQOTRGL-----GCITSLTG 195  
 Db 247 -----PSRLHYRNQWRSVTSVGAEEATSGLVMTCHISLVNSTYN 290  
 QY 196 -----RKNQYGEVQVYSTATQSEL 216  
 Db 291 TPYTGALGLDPALEFRNLTPGNTNTRVRSSTARRLRGADGAEITTAATFEM 350  
 QY 217 A-----TCVNGV-----CMTVEH-----GAG-----SKTLAEPKG--PIT 244  
 Db 351 KDLVFTSTNGVEIGRGALTLFNLADTLGLPTELISAGGQLFYSRPVVSANGEPV 410  
 QY 245 QMYTNVDDIDVGMQAPPGARSMPTCTCGSSDLYLV---TRADYIPVRRRDSRG--SLS 300  
 Db 411 KLYTSVENA-----QDDKGIAPPHDIDGESRVVIGDYNGHQRPTSPAPSRFVTLR 466  
 QY 301 PRPVSYLK-----GSSGPIICPSGHAAGIFRAAVCTRGVAKAVDFIPV 344  
 Db 467 ANDVLMISLTAEXDOSTYSGSTGPVYV--SDSVTLVNVATGAQAVASLDMTKY 519

## RESULT 15

H71426  
 hypothetical protein - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 A/Variety: Columbia  
 C/Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
 C/Accession: H71426  
 R/Beyan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
 P.; Wedler, H.; Medler, E.; Wandt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giehl  
 vanagh, T.; Hempel, S.; Kotter, P.; Shitlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
 Nature 391, 485-488, 1998  
 A/Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Paigdemenech,  
 ehoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reehman, S.; Ansc  
 C.; Chaitazis, N.  
 A/Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
 A/Reference number: A71400; MUID:98121113; PMID:9461215  
 A/Accession: H71426  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-692 <BEV>  
 A/Cross-references: GB:Z97340; NID:g2244950; PID:e327492; PID:g2244965  
 C/Genetics:  
 A/Map position: 4COP9-4G3845

Query Match 5.6%; Score 102.5; DB 2; Length 692;  
 Best Local Similarity 23.9%; Pred. No. 1.2;  
 Matches 91; Conservative 41; Mismatches 127; Indels 121; Gaps 22;

QY 8 PPLNVRGGRPAIILTCVAVPELIDITKLLAFGIMVTLQAGITVYV--FVRAGLI 65  
 Db 101 PTVASVYGSNCIE---KEPEIIMDTLELR--FLTLQWL---FSKPPAVLESIGYT 151  
 QY 66 PACMLVR--FAAGHYVQVAFM-----KLAALGTGVYDHLTP 101  
 Db 152 HEDVLLQPRAGVGHIMQAPAFITIRDTNSKILLIRGTHSINKDTLTAAGAVVPFHHSV 211





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.34864 Seconds

(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-13  
Perfect score: 1842

Sequence: 1 AHQVWIPINVRGRDAII.....RGVAKAVDFIPVSEMTTMR 352

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1773	96.3	3010	1	P26652 h genome po
2	1766	95.9	3010	1	P00259 h genome po
3	1749	95.0	3010	1	P29846 h genome po
4	1699	92.2	3010	1	P26653 h genome po
5	1616	87.7	3011	1	P26664 h genome po
6	1605	87.1	3011	1	P27958 h genome po
7	1303	70.7	3033	1	P26661 h genome po
8	1301	70.6	3033	1	P26660 h genome po
9	102.5	5.6	660	1	P29326 hepatitis e
10	102.5	5.6	660	1	P33426 hepatitis e
11	101	5.5	564	1	P37107 arabisdopsis
12	101	5.5	600	1	P33611 mus musculi
13	93.5	5.1	1380	1	C27675 leishmania
14	93	5.0	434	1	Q27675 leishmania
15	92.5	5.0	706	1	P27425 equus caball
16	92	5.0	659	1	Q00270 hepatitis e
17	91.5	5.0	485	1	Q04611 hepatitis e
18	91.5	5.0	660	1	P14336 t genome po
19	91.5	5.0	3414	1	Q01299 t genome po
20	90.5	4.9	3414	1	Q01299 t genome po
21	89.5	4.9	961	1	Q8287 yersinia pe
22	87.5	4.8	3412	1	P07770 t genome po
23	86.5	4.7	347	1	P46487 eucalyptus
24	86.5	4.7	470	1	P05883 influenza a
25	86	4.7	338	1	Q05026 neisseria a
26	86	4.7	433	1	Q34255 wolnietia s
27	85	4.6	470	1	P03472 influenza a
28	85	4.6	730	1	Q8287 methanocarc
29	85	4.6	854	1	PMP2 SCHPO
30	85	4.6	3313	1	Q8287 schizosacch
31	84.5	4.6	1705	1	P70289 mus musculi
32	84	4.6	309	1	P56500 ratius norv
33	84	4.6	339	1	Q81P0 corynebacte

34	84	4.6	423	1	NR09 MOUSE	Q9vrt6 mus musculi
35	84	4.6	470	1	NR09 MOUSE	P31510 influenza a
36	84	4.6	503	1	YA02 MYCTU	O05586 mycobacteri
37	83.5	4.5	538	1	DAC ACTSP	P39045 actinomadar
38	83.5	4.5	1399	1	RPOC PSEAE	Q9hwc9 pseudomonas
39	83	4.5	341	1	MDHW BRANA	Q43744 brassica na
40	82.5	4.5	453	1	NR09 BRANA	P03470 influenza a
41	82.5	4.5	711	1	H2PA RHIME	Q8vrt3 rhizobium m
42	82.5	4.5	1022	1	CA26 CHICK	P15988 gallus gall
43	82	4.5	309	1	UCP2 MOUSE	P70406 mus musculi
44	82	4.5	384	1	MK08 BRARE	Q9dgd9 brachydanio
45	82	4.5	403	1	PGK CHIMU	Q9pin4 chlamydia m

## ALIGNMENTS

```

RESULT 1
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26652;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP66) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (Isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL; D90208; BAA14233.1; -

```

QY	1	AHQQWTPPLANNGGSDAIIILLCAAPHELIIPITKLLIAEPGLWVLQAGTKVPFVR	60
		Best Local Similarity 94.6%; Pred.No. 3e-141; Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;	
Db	855	AHLQWVPPPLNVGGRDAIILLCAVAPHELIIPITKLLIAEPGLWVLQAGTKVPFVR	914
QY	61	AAGLIRACMLVRRKAAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAEVP	120
Db	915	AAGLIRACMLVRRKAAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAEVP	974
QY	121	IFSDMEVKITTWGADTPAACDIIISGLPVARRRREILLGANDPBEQGMPLLAPITAYSQ	180
Db	975	VFSMEKMLITWGAADTPAACDIIISGLPVARRRREILLGADPBEQGMPLLAPITAYSQ	1034
QY	181	QTRGLACITLSLTGRDXNOVEGEVQVSPATOSPLATCNGVCMTVFEHAGSKTLIAGPK	240
Db	1035	QTRGLACITLSLTGRDXNOVEGEVQVSPATOSPLATCNGVCMTVFEHAGSKTLIAGPK	1094
QY	241	GPITQMTYNVDQDILVGMQAPPGARSMTPTCGSSDLYLVTRHADVLPVRRRGRSGILS	300
Db	1095	GPITQMTYNVDQDILVGMQAPPGARSMTPTCGSSDLYLVTRHADVLPVRRRGRSGILS	1154
QY	301	PRPSTYKSGSGGGLPCPSGHAIVGIPPAACVTFGAKAVAFIPESMETMR	352
Db	1155	PRPSTYKSGSGGGLPCPSGHAIVGIPPAACVTFGAKAVAFIPESMETMR	1206

RESULT 2

POLG\_HCVJT

ID	POLG_HCVJT	STANDARD;	PRT;	3010 AA.
AC	Q00269;			
AD	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Genome polypeptide [Contans: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP25); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7: Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadvirid) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate HC-Jr) (HCV).			
OC	Vireuses; ssRNA positive-strand vireuses, no DNA stage; Flaviviridae; Hepacivirus.			
CC	NCBI_taxid=31642;			
CC	[1]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=92295714; PubMed=1318627;			
RX	Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shinocho K.;			
RA	"Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals."			
RT	Virus Res. 23:39-53(1992).			
RL	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).			
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way			

FT	CARBOHYD	22240	2240	N-LINKED (GLCNAC. . .)	(POTENTIAL)
FT	CARBOHYD	2529	2529	N-LINKED (GLCNAC. . .) <td>(POTENTIAL)</td>	(POTENTIAL)
FT	CARBOHYD	2788	2788	N-LINKED (GLCNAC. . .) <td>(POTENTIAL)</td>	(POTENTIAL)
SEQ	SEQUENCE	3010 AA;	326573 MW;	94A1C77435D642B8	CRC64;

  

Query Match	Best Local Similarity	95.9%;	Score 1766;	DB 1;	Length 3010;				
Matches	334;	Conservative	7;	Mismatches	11;	Indels	0;	Gaps	0;
QY	1	AHLQVMIPLINVEGRDAIILLTCVAPEHIFDITKLLLAIFGPIAMVLAQGITKPYEVR	60						
Db	855	AHLQVMIPLINVEGRDAIILLTCVAPEHIFDITKLLLAIFGPIAMVLAQGITKPYEVR	914						
QY	61	AQGLIRKCMVLRKAGSHYQMAFMKLAITGTYVTHLPLOMAHAGRLDAVAEVP	120						
Db	915	AQGLIRKCMVLRKAGSHYQMAFMKLAITGTYVTHLPLOMAHAGRLDAVAEVP	974						
QY	121	IFSDMEKIIITWGADTACDIIISGLPVSRARGREILIGADNDEGGWMLPITAYSQ	180						
Db	975	VFSDEMEKIIITWGADTACDIIISGLPVSRARGREILIGADNDEGGWMLPITAYSQ	1034						
QY	181	QTRGLCCIIITSLTGRDKNVESEVQVSTATOSFLATCVNGVMTVPHGAGSKTLGPK	240						
Db	1035	QTRGLCCIIITSLTGRDKNVESEVQVSTATOSFLATCVNGVMTVPHGAGSKTLGPK	1094						
QY	241	GPIIOMTANVDQDLVGMQAPRGARSMPCGSSDLVLRHADVIVRRRGDSRGLLS	300						
Db	1095	GPIIOMTANVDQDLVGMQAPRGARSLPCTCGSSDLVLRHADVIVRRRGDSRGLLS	1154						
QY	301	PRPVSYLKGSSGGPILCPSGHAGVIFRAAICTRGVAKAVDFIVESMETMR	352						
Db	1155	PRPVSYLKGSSGGPILCPSGHAGVIFRAAICTRGVAKAVDFIVESMETMR	1206						

  

RESULT 3	POLG_HCVTW	STANDARD;	PRT;	3010 AA.
AC	P29846;			
DT	01-APR-1993 (Ref. 25, Created)			
DT	01-APR-1993 (Ref. 25, Last sequence update)			
DT	10-OCT-2003 (Ref. 42, Last annotation update)			
DE	Genome polypeptide [contains: capsid protein C (core protein) (P22);			
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2			
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)			
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)			
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein			
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein			
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.45)].			
OS	Hepatitis C virus (isolate Taiwan) (HCV).			
OC	Virusess; ssRNA positive-strand viruses; no DNA stage; Flaviviridae;			
OC	Hepadivirus.			
OC	NCBI_TaxID=31645;			
OC	[1]_			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92230206; PubMed=1314449;			
RA	Chen P.U., Lin M.H., Tai K.F., Liu P.C., Lin C.U., Chen D.S.;			
RA	"The Taiwanese hepatitis C virus genome: sequence determination and			
RT	mapping the 5' terminus of viral genomic and antigenomic RNA.";			
RT	Virology 188:102-113(1992).			
CC	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are			
CC	hydrophobic, suggesting a possible membrane-related function. NS3			
CC	and NS5 may play a role in the viral RNA replication.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral			
CC	precursor polypeptide, commonly with Asp or Glu in the P6			
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +			
CC	[RNA] (N).			
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a			
CC	lipoprotein envelope. The envelope consists of two proteins:			
CC	protein M and glycoprotein E. The nucleocapsid is a complex of			
CC	protein C and RNA.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL, M64754; -, NOT\_ANNOTATED\_CDS.

PIR, A40244; GKMVTV.

PDB, 1N64; 25-FEB-03.

PDB, 1NS3; 08-APR-98.

MEROPS; S29.001; -.

MEROPS; U39.001; -.

InterPro; IPR009003; Cys\_Ser\_trypsin.

InterPro; IPR001410; DEAD.

InterPro; IPR002522; HCV\_capsid.

InterPro; IPR002521; HCV\_core.

InterPro; IPR002519; HCV\_env.

InterPro; IPR002518; HCV\_NS1.

InterPro; IPR002517; HCV\_NS2.

InterPro; IPR002516; HCV\_NS4.

InterPro; IPR002515; HCV\_NS4B.

InterPro; IPR002514; HCV\_NS4C.

InterPro; IPR002513; HCV\_NS4D.

InterPro; IPR002512; HCV\_NS4E.

InterPro; IPR002511; HCV\_NS4F.

InterPro; IPR002510; HCV\_NS4G.

InterPro; IPR002509; HCV\_NS4H.

InterPro; IPR002508; HCV\_NS4I.

InterPro; IPR002507; HCV\_NS4J.

InterPro; IPR002506; HCV\_NS4K.

InterPro; IPR002505; HCV\_NS4L.

InterPro; IPR002504; HCV\_NS4M.

InterPro; IPR002503; HCV\_NS4N.

InterPro; IPR002502; HCV\_NS4O.

InterPro; IPR002501; HCV\_NS4P.

InterPro; IPR002500; HCV\_NS4Q.

InterPro; IPR002499; HCV\_NS4R.

InterPro; IPR002498; HCV\_NS4S.

InterPro; IPR002497; HCV\_NS4T.

InterPro; IPR002496; HCV\_NS4U.

InterPro; IPR002495; HCV\_NS4V.

InterPro; IPR002494; HCV\_NS4W.

InterPro; IPR002493; HCV\_NS4X.

InterPro; IPR002492; HCV\_NS4Y.

InterPro; IPR002491; HCV\_NS4Z.

InterPro; IPR002490; HCV\_NS4A.

InterPro; IPR002489; HCV\_NS4B.

InterPro; IPR002488; HCV\_NS4C.

InterPro; IPR002487; HCV\_NS4D.

InterPro; IPR002486; HCV\_NS4E.

InterPro; IPR002485; HCV\_NS4F.

InterPro; IPR002484; HCV\_NS4G.

InterPro; IPR002483; HCV\_NS4H.

InterPro; IPR002482; HCV\_NS4I.

InterPro; IPR002481; HCV\_NS4J.

InterPro; IPR002480; HCV\_NS4K.

InterPro; IPR002479; HCV\_NS4L.

InterPro; IPR002478; HCV\_NS4M.

InterPro; IPR002477; HCV\_NS4N.

InterPro; IPR002476; HCV\_NS4O.

InterPro; IPR002475; HCV\_NS4P.

InterPro; IPR002474; HCV\_NS4Q.

InterPro; IPR002473; HCV\_NS4R.

InterPro; IPR002472; HCV\_NS4S.

InterPro; IPR002471; HCV\_NS4T.

InterPro; IPR002470; HCV\_NS4U.

InterPro; IPR002469; HCV\_NS4V.

InterPro; IPR002468; HCV\_NS4W.

InterPro; IPR002467; HCV\_NS4X.

InterPro; IPR002466; HCV\_NS4Y.

InterPro; IPR002465; HCV\_NS4Z.

InterPro; IPR002464; HCV\_NS4A.

InterPro; IPR002463; HCV\_NS4B.

InterPro; IPR002462; HCV\_NS4C.

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDE215 CRC64;

Query Match 95.0%; Score 1749; DB 1; Length 3010;  
Best Local Similarity 92.98; Pred. No. 3.2e-139; Indels 0; Gaps 0;  
Matches 327; Conservative 14; Mismatches 11;

QY 1 AHLQWIPPLNVRGGRDAIILLTCVAPBELIFDTKLLAIFGLVNLQGITKPYFVR 60  
DB 955 AHLQWIPPLNVRGGRDAIILLTCVAPBELIFDTKLLAIFGLVNLQGITKPYFVR 914  
QY 61 AGLIRACMLVRKAAGHYQYQAMFKLALGTGVYDHLPLQDMAHAGRLDAVAEPV 120  
DB 915 AGLIRACMLVRKAAGHYQYQAMFKLALGTGVYDHLPLQDMAHAGRLDAVAEPV 974  
QY 121 IFSMEVYKILTWGADTAACGDIISGLPVSARSGEILLGPADNFEQGWRLAPITAYSQ 180  
DB 975 VFSDEMTKITWGDITACGDIISGLPVSARSGEILLGPADNFEQGWRLAPITAYSQ 1034  
QY 181 QTRGLGCTITSLTRDQNVQEGVQVSTRTQSLATCNQGVNTYFPHAGSKTLGAPK 240  
DB 1035 QTRGLGCTITSLTRDQNVQEGVQVSTRTQSLATCNQGVNTYFPHAGSKTLGAPK 1094  
QY 241 GPITQMTNVVODLVGMQAPRGASMTPTCGSSDLVLRHADVIFVRRGDSRGLS 300  
DB 1095 GPITQMTNVVODLVGMQAPRGASMTPTCGSSDLVLRHADVIFVRRGDSRGLS 1154  
QY 301 PRPVSYLNGSSGGLPLCPGSHAVGIFRAVCTRGAVAVDFIYVSMETTWK 352  
DB 1155 PRPVSYLNGSSGGLPLCPGSHAVGIFRAVCTRGAVAVDFIYVSMETTWK 1206

RESULT 4  
ID POLG HCVRK STANDARD; PRT; 3010 AA.  
AC P26653;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
DE Hepatitis C virus (isolate BK) (HCV).  
OS Hepatitis C virus  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
OC Hepacivirus.  
OX NCBI\_Taxid=11105;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91140698; PubMed=1847440;  
RA Takamiizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;  
RT "Structure and organization of the hepatitis C virus genome isolated from human carriers";  
RL J. Virol. 65:1105-1113 (1991).  
RN [2]  
RP SEQUENCE OF 1487-1500.  
RX MEDLINE=96325224; PubMed=8647104;  
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornatecky L.;  
RT "Non-structural protein 3 of hepatitis C virus inhibits





```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 87.7%; Score 1616; DB 1; Length 3011;
Best Local Similarity 84.1%; Pred. No. 5,5e-128;
Matches 296; Conservative 28; Mismatches 28; Indels 0; Gaps 0;

QY 1 ARLQVWIPLVNVRGGRDAIILTCVAHPELIPDITKLLALFGLPLVLAQGITKVPYFVR 60
DB 855 AQLHWIPIPLNVGRGRDAVILLMCVAHPTLFDITKLLAVFGPLWTLQASLTKVPYFVR 914
QY 61 AAGLIRACMLVRKAGGVYQVAFMKLAALGTGVYDHLPLDQMAHAGLRDLAVAVEPV 120
DB 915 VQGLRFLCARLARKWIGVHVMYITKGLALGTIVYVHLPLDQMAHAGLRDLAVAVEPV 974
QY 121 IFSDFMEKIIITWGAADTAACDITLSGLPVSARGRSEILLGADNFEQGMWLLAPITAYSQ 180
DB 975 VFGSEMERKLTWGDATACGDIINGLPLVSARGRSEILLGADNFEQGMWLLAPITAYSQ 1034
QY 181 QTRRLCCITLTLGRPKNVEGEVQVSTATGSEFLATCVNGCVMTYFHASGSKITLAGPK 240
DB 1035 QTRRLCCITLTLGRPKNVEGEVQVSTATGSEFLATCVNGCVMTYFHASGSKITLAGPK 1094
QY 241 GPIPTWNTVDQDLVGMQAPGARSMTPCGSSDLVLYTRHADVIFVRRCDSRGSILS 300
DB 1095 GPIPTWNTVDQDLVGMQAPGARSMTPCGSSDLVLYTRHADVIFVRRCDSRGSILS 1154
QY 301 PRPVSYIKSGSSGGLPLCPSGHAGVIFPAVCTRGVAVADFIPEVSEMTTMR 352
DB 1155 PRPVSYIKSGSSGGLPLCPSGHAGVIFPAVCTRGVAVADFIPEVSEMTTMR 1206

RESULT 6
POLG_HCVH STANDARD; FRT; 3011 AA.
ID POLG_HCVH
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchausti G., Zebede S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RL virus: comparison with American and Japanese isolates.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

```

```

RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=96154321; PubMed=9493270;
RA Kim J.U., Morsemstein K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA5534.1; -
DR PIR; A36814; GNVVCH.
DR PDB; 1AIV; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -
DR MEROPS; U39.001; -
DR TRANSFAC; T04155; -
DR INTERPRO; IPR009003; Cys_ser_trypsin.
DR INTERPRO; IPR001410; DEAD.
DR INTERPRO; IPR002522; HCV capsid.
DR INTERPRO; IPR002521; HCV env.
DR INTERPRO; IPR002519; HCV env.
DR INTERPRO; IPR002511; HCV NS1.
DR INTERPRO; IPR002518; HCV NS2.
DR INTERPRO; IPR000745; HCV NS4A.
DR INTERPRO; IPR001490; HCV NS4B.
DR INTERPRO; IPR002868; HCV NS5A.
DR INTERPRO; IPR002166; HCV RdRp.
DR INTERPRO; IPR001650; Helicase_C.
DR INTERPRO; IPR004109; Peptidase_C29.
DR INTERPRO; IPR007094; RNA_pol_PS.
DR INTERPRO; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01558; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.

```

Query Match	Best Local Similarity	87.1%;	Score 1605;	DB 1;	Length 3011;
Matches 294;	Conservative 30;	Mismatches 28;	Indels 0;	Gaps 0;	







01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Burma) (HEV).  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 CC Hepatitis E-like viruses.  
 NCBI\_TaxID=31767;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92024067; PubMed=1926770;  
 RA Tem A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,  
 RA Fry K.E., Reyes G.R.,; molecular cloning and sequencing of the  
 RT "Hepatitis E virus (HEV): full-length viral genome."  
 RL Virology 185:1120-131(1991).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; M73218; AAA45736.1; -  
 DR PIR; C40778; VHWMH2.  
 DR InterPro; IPR004261; SP2.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF03014; SP2; 1.  
 KW Signal.  
 FT CHAIN  
 FT SIGNAL 1 19  
 FT CHAIN 20 660  
 SQ SEQUENCE 660 AA; 70978 MW; 58322A013CCCA61C CRC64;  
 -----  
 Query Match 5.6%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 0.59; Indels 151; Gaps 19;  
 Matches 82; Conservative 52; Mismatches 130;  
 QY 35 TKLL--AIFGLMVLQAG-----ITKPYFVR--AAGLRACMLYKAGGHVQMA 83  
 DB 151 TNLVLAAPLSPLPLQDGTNTHIMATEASNAQVRVATIRYRPLVPAVAGTAISIS 210  
 QY 84 FMKLAALGTGYVYHLPLODMWAHGLDLAVAEVPIFSDMEVKIITWADTAACGDI 143  
 DB 211 FWPQTITPTTSV-----DMNSITSDVRILVPGIASLVI----- 246  
 QY 144 SGLPVSARRGREILLGPAD--NFEQGWRLLAPI--TAVSOQTRGL-----GCITISLNG 195  
 DB 247 -----PSERLHYRNQGRSVETSGVAEEETSLVWLCHGSLVNSYTN 290  
 QY 196 -----RDXQVBEQVQVVSATQSEFL 216  
 DB 291 TPYTGALGLDPALEFRLNLTPTNTVRSYSTARHRLRGADGTALTTTAATRM 350  
 QY 217 A---TCVNGV-----CMTVFR-----GAG-----SKTLAEPKG--PIT 244  
 DB 351 KDLFTSTNGVGBIGRGIALTLFNLADTLGLPTELISSAGGLFYSRPVVSANGEPV 410  
 QY 245 QMTYTNVDQLVGMQAPPGARSMTPTCGSSDLVYV---TRHADVIYVRRGDSRG--SLIS 300  
 DB 411 KLYTSVENA---QDQKGLAIPHDIDLGESRVIQDYDNQCHQDRPTSPAPSRPFSVLR 466  
 QY 301 PRPVSYLK-----GSSGGLLCPSGHVAIGFRAAVCTRGVAKAVDFIV 344  
 DB 467 ANDVLMLSLTAAEVDQSTYGSSTGFPYV--SDSVTLVNVATGAQAVARSLDWTYV 519

01-FEB-1994 (Rel. 28, Created)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Burma) (HEV).  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 CC Hepatitis E-like viruses.  
 NCBI\_TaxID=31774;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92115700; PubMed=1731327;  
 RA Tearev S.A., Emerson S.U., Reyes G.R., Teareva T.S., Legters L.J.,  
 RA Malik I.A., Iqbal M., Purcell R.H.,; "Characterization of a prototype strain of hepatitis E virus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; M80581; AAA45727.1; -  
 DR InterPro; IPR004261; SP2.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF03014; SP2; 1.  
 KW Signal.  
 FT CHAIN  
 FT SIGNAL 1 22  
 FT CHAIN 23 660  
 SQ SEQUENCE 660 AA; 70980 MW; 8083BC3C7B46FD3 CRC64;  
 -----  
 Query Match 5.6%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 0.59; Indels 151; Gaps 20;  
 Matches 82; Conservative 54; Mismatches 126;  
 QY 35 TKLL--AIFGLMVLQAG-----ITKPYFVR--AAGLRACMLYKAGGHVQMA 83  
 DB 151 TNLVLAAPLSPLPLQDGTNTHIMATEASNAQVRVATIRYRPLVPAVAGTAISIS 210  
 QY 84 FMKLAALGTGYVYHLPLODMWAHGLDLAVAEVPIFSDMEVKIITWADTAACGDI 143  
 DB 211 FWPQTITPTTSV-----DMNSITSDVRILVPGIASLVI----- 246  
 QY 144 SGLPVSARRGREILLGPAD--NFEQGWRLLAPI--TAVSOQTRGL-----GCITISLNG 195  
 DB 247 -----PSERLHYRNQGRSVETSGVAEEETSLVWLCHGSLVNSYTN 290  
 QY 181 QT-RGLLGI-----ITSLGRDKNQ-----VEGEQVVSATQSEFL 216  
 DB 291 TPYTGALGLDPALEFRLNLTPTNTVRSYSTARHRLRGADGTALTTTAATRM 350  
 QY 217 A---TCVNGV-----CMTVFR-----GAG-----SKTLAEPKG--PIT 244  
 DB 351 KDLFTSTNGVGBIGRGIALTLFNLADTLGLPTELISSAGGLFYSRPVVSANGEPV 410  
 QY 245 QMTYTNVDQLVGMQAPPGARSMTPTCGSSDLVYV---TRHADVIYVRRGDSRG--SLIS 300  
 DB 411 KLYTSVENA---QDQKGLAIPHDIDLGESRVIQDYDNQCHQDRPTSPAPSRPFSVLR 466  
 QY 301 PRPVSYLK-----GSSGGLLCPSGHVAIGFRAAVCTRGVAKAVDFIV 344  
 DB 467 ANDVLMLSLTAAEVDQSTYGSSTGFPYV--SDSVTLVNVATGAQAVARSLDWTYV 519

RESULT 10  
 VST2\_HEVPA STANDARD; PRT; 660 AA.  
 AC P33426;

RESULT 11  
 SRSC\_ARATH STANDARD; PRT; 564 AA.  
 AC P3107; O82570;



CC alpha/primase complex to the cellular replication machinery (By  
 CC similarity).  
 CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme  
 CC (subunits A, B, C and D), which is assembled throughout the cell  
 CC cycle. The largest subunit (subunit A) has DNA polymerase  
 CC activity, the two smallest subunits (subunits C and D) have DNA  
 CC primase activity. Subunit B binds to subunit A.  
 CC SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M  
 CC -1- PHASE (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D13546; BA02746.1; -  
 CC PIR: B46642; B46642.  
 CC MGD: MGI:96690; P01A2.  
 CC InterPro: IPR007200; DNA\_pol\_alpha\_B.  
 CC Pfam: PF04058; DNA\_pol\_alpha\_B.1.  
 CC DNA replication; Nuclear protein; Phosphorylation.  
 CC DOMAIN 101 107 POLY-GU.  
 CC FT 115 157 PRO/SER/THR-RICH (HYDROPHILIC).  
 CC SEQUENCE 600 AA; 66267 MW; 79F9ABE6EF33FEEC CRC64;  
 SQ  
 Query Match 5.5%; Score 101; DB 1; Length 600;  
 Best Local Similarity 24.8%; Pred. No. 0.71;  
 Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;  
 QY 64 LIRACMVRKRAAGHYVOM-AFMKLAALT-----GTYYVDH-----TPLDMA 106  
 Db 27 LAECVIRQEDSMVSELINFTCSAGKTCITVDILNSFEVILANKLSTAMHSASDSC 86  
 QY 107 HAGRLDAVAVEPIFSDMEVKIITWGADPAACGDI--ISGLP-----VSARGREI 156  
 Db 87 HAGRDI-VSIQELIEAEEHEETLLSYTPSKGFLKRVSTPTPTKRSVARSPPQ- 144  
 QY 157 LIGPADNPEGGAMFLAPITAYSCQTKLGCITTSITGDKXVBEQVNSTANQSL 216  
 Db 145 LLSFSS-----FSPSATPSOK-----YTSRTNR-----DEVVTTGSAQ- 178  
 QY 217 ATCVNGVCMVTFHGAGSKTL--AGPKGPIITOMTYNVQDILVG 256  
 Db 179 -----GLSMGSGRGSGSVSLKVVDPRLTGSYKAMQQLMG 215  
 RESULT 13  
 CYAA LEIDO STANDARD; PRT; 1380 AA.  
 ID CYAA LEIDO STANDARD; PRT; 1380 AA.  
 AC Q27675;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Receptor-type adenylylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-  
 DE lyase) (Adenylyl cyclase).  
 GN RAC-A.  
 OS Leishmania donovani.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1S Sudanese;  
 RX MEDLINE=95340554; PubMed=7615561;  
 RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.,  
 RT "A family of putative receptor-adenylylate cyclases from Leishmania  
 RT donovani".  
 RL J. Biol. Chem. 270:17551-17558(1995).  
 CC -1- FUNCTION: Could act as a receptor for a unknown ligand.

CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -1- COPACITOR: Binds 1 magnesium ion per subunit (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
 CC -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (premacigote)  
 CC but not in the mammalian host stage of the parasite life cycle.  
 CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U17042; AAA74998.1; -  
 CC PIR: T18309; T18309.  
 CC InterPro: IPR001054; G\_cyclase.  
 CC Pfam: PF00211; guanylate\_cyc; 1.  
 CC SMART: SM00044; CYCC; 1.  
 CC DR PROSITE: PS50125; GUANYLATE CYCLASES\_2; 1.  
 CC DR Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein;  
 CC Metal-binding; Magnesium.  
 CC DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 35 55 POTENTIAL.  
 CC DOMAIN 56 891 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 892 912 POTENTIAL.  
 CC DOMAIN 913 1380 CYTOPLASMIC (POTENTIAL).  
 CC METAL 938 938 MAGNESIUM (BY SIMILARITY).  
 CC METAL 981 981 MAGNESIUM (BY SIMILARITY).  
 CC CARBOHYD 422 422 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 478 478 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 497 497 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 1380 AA; 151692 MW; 6B2DF7D3C1107A0 CRC64;  
 Query Match 5.1%; Score 93.5; DB 1; Length 1380;  
 Best Local Similarity 21.0%; Pred. No. 8.4;  
 Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;  
 QY 14 GGRDALILLTCAVHP---ELIFDIYTKL-----LIALFGPLM-----V 47  
 Db 101 GGRPIKITL---HPDPCDNLVDAEVLHSLAQEKILAVIGYLDRLTALPALSNADV 155  
 QY 48 LQAGITKVP-----YFVRAGLIRACMIVRKAAGHYV-----QVAFPK 86  
 Db 156 YQSGMLLILPFTGSSGVRTWSGVYFTRAEPMVELKVLVLM-----HIVRLARRVAFMR 210  
 QY 87 LIALTG-----TYYVDLTPLODMAHAGRLDAVA-VEPIFSDMEVK----- 128  
 Db 211 ---LTGMHFGGELTYVQDTLSL-----LRDPALVLTVPYSESVVEDEAPFAMAD 260  
 QY 129 -----ITWGADPAACGDIISGLIPYSARGREILG-----PANDFEGQ 167  
 Db 261 TNPQVITWAAAPQVQYVLEKVLTDPTSSAVVISCMIQRVEDVYRLLISAGIKFQ 320  
 QY 168 GWRLLAPITAYSCQTKLGCITTSITGDKXVBEQVNSTANQSL 223  
 Db 321 DGRILASATL-----SPVSEGLKYNVEVLKAMQSNVIE---NSG 356  
 QY 224 CWTVEFGAGSKTLAGEK---GITOMTYNVQDILVGMQAPPGARSMTECTGSSLLYLIV 279  
 Db 357 SPDYPPDDSTETLGRKARSEAPLSRKYT-VDE---FQAHPISIAKLMALGMLSTLVQ 412  
 QY 280 TRHADVIYPRRRDSDGSLSPPV---SYLKSSGGPLIC-PSGAVGIFRAVCTRGV 335  
 Db 413 TLEQTMVYKSTYKAGLGNQRNFVIGSDVVLGDYGGP---CEFLAFLG--ASCYQNGG 468  
 QY 336 AKAV-----DFIVESME 348  
 Db 469 HSSILTVLQVASWDIVPDSSEK 490

```

RESULT 14
TOLB_CHLITE STANDARD; PRT; 434 AA.
ID TOLB_CHLITE
AC Q8RECO:
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE TOLB protein precursor.
GN TOLB OR C70636.
OS Chlorobium tepidum.
CC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
CC Chlorobium.
CC NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TL5 / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vanatvehan J., Khoult H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC -1- FUNCTION: Involved in the tomb-independent uptake of proteins (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: Belongs to the tolb family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AEO12837; AAM71875.1; -.
DR TIGR; CT0636; -.
DR HAMAP; MF 00671; -.
KM Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 434 TOLB protein.
SQ SEQUENCE 434 AA; 47275 MW; F1A347B89C7A0F99 CRC64;

Query Match 5.0%; Score 93; DB 1; Length 434;
Best Local Similarity 21.2%; Pred. No. 2.3;
Matches 69; Conservative 47; Mismatches 139; Indels 70; Gaps 14;

QY 42 FGLMLVQAGITVPPFVRAGLIRACMLVRKAAAGHYQAMFKALALTYVDHLTP 101
DB 8 FACLCTMLFGLMFLVPPFLRAE-VGEYIARF-EGASRIAVLDKTSADGGKQ----- 58
QY 102 LQDMAHA-----GLDLAAVEVEFVPSDMVEKLIITGADTAACGDIISGLPVS 149
DB 59 -REMASSLDVTINKGIDFTGLFHLRA--PLNIRNQNGNINPASIAGGDIYAGGSTY 115
QY 150 ARGRERIL-----LQPADNEFGGCMRLIAPTVASQOTRGLGCTITS 192
DB 116 KRSGREPLEMHWYSSGSKSLARTYTGESQRLAIGRCADLVELLTGKESVGFRIYV 175
QY 193 LTRGRDKNO-----VEGEVYVSTATOSFLATCV-----NGVCWYVHGAAGKTLAAGPG 241
DB 176 VANRTQKEIYMDPFGEENVVQLTNSRSISLTPAVSPDGTIAMDY--TSGKPNLYIKN 233
QY 242 PITQMTNVQDQIV---GWQAPPGARSMTPTCT--GSSDLVLYVRHADVIFVRERGSR 295
DB 234 IATGAKVSVKRGVCISPAWR--PGTNTLVYTLISYEGDQDLYLI--RADGTVERRLTKGG 289
QY 296 GSLLSP-----RPVSYLKSGSSGCP 314

```

```

DB 290 GIDVSPFSPDGSXMAFVSTRQSCP 314

RESULT 15
TREE_HORSE STANDARD; PRT; 706 AA.
ID TREE_HORSE
AC P27425;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sero transferrin precursor (transferrin) (Siderophilin) (Beta-1-metal
DE binding globulin).
GN TF.
OS Equus caballus (Horse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OS Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
CC NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277958; PubMed=8504171;
RA Carpenter M.A., Broad T.E.;
RT "The cDNA sequence of horse transferrin.";
RL Biochim. Biophys. Acta 1173:230-232(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraembryonic tissue;
RA McDowell K.J., Adams M.H., Baker C.B.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Transferrins are iron binding transport proteins which
CC can bind two atoms of ferric iron in association with the binding
CC of an anion, usually bicarbonate. It is responsible for the
CC transport of iron from sites of absorption and heme degradation to
CC those of storage and utilization. Serum transferrin may also have
CC a further role in stimulating cell proliferation.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -1- DOMAIN: Composed of two homologous domains.
CC -1- SIMILARITY: Belongs to the transferrin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M69020; AAA30958.1; -.
DR EMBL; U21127; AAA63684.1; -.
DR PIR; S33761; S33761.
DR HSRF; P02787; IABE.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PRO0422; TRANSFERRIN.
DR SMART; SMO0094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KM Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
Signal.
FT SIGNAL 1 19
FT CHAIN 20 706
FT REPEAT 20 357
FT REPEAT 358 706
FT DISULFD 26 64
FT DISULFD 36 55
FT DISULFD 134 215
FT DISULFD 174 190
FT DISULFD 177 198
FT DISULFD 187 200
FT DISULFD 248 262

```

```
FT DISULFID 360 623 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 376 389 BY SIMILARITY.
FT DISULFID 423 701 BY SIMILARITY.
FT DISULFID 441 664 BY SIMILARITY.
FT DISULFID 474 550 BY SIMILARITY.
FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;
```

```
Query Match 5.0%; Score 92.5; DB 1; Length 706;
Best Local Similarity 21.5%; Pred. No. 4.5;
Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

QY 82 MAEKKLALNGTYV---YDHLPQDMAHGLRDLAVAVEPIFSDMEVKITWGA----134
Db 321 LGFLRIPFAMDWTWLYGTYVT-----AIRNLREDIRPEVPKD-ECKKVMCAIGHH 371
QY 135 DTAACGD-IISGLPVASARRGR-----ELLGPDNFEQGQWRL-----LAPITAY 178
Db 372 EKVKCDMSVNSGNIICESASTEDCIAXIKGEADMSLDGGFIYIAGKCGIPLVLAIE 431
QY 179 SQCTRGILGCIITSLTRDKNQVGEVQVSTATQSFLATCVNGCWTVPFHGASKTLAG 238
Db 432 NYETRSSACVDTPPEGTH-----AIVAVKSSSDPDLT-----N-----NSLKG 470
QY 239 PKGPIITOMYTNVDODLVGMQAPPGARSMTPCTCGSSDLYLVTRRADVIPIVRRGDSRSL 298
Db 471 KK-----SCHTGVDR-TAGWNIPMGL-----LYSEIKHCEFDKFFREGCARGYR 513
QY 299 LSPRPVSLTKSSSGGP-LLC-PSGHA-----VGIFRAVCTRGVAKAVDFIPVESME 348
Db 514 RNSLTCLNLCIGSAGSGPGECEPNNHERYGYGTGAFCLVEKGDVA---FVKQQTVE 566
```

Search completed: May 6, 2004, 09:31:53  
Job time : 8.34864 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 32.9219 Seconds  
(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-13  
Perfect score: 1842  
Sequence: 1 AHQVITPPUNVGRDAII.....RGVAKAVDPFVPSMETTKR 352

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phase:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriaph:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	97.3	3010	12 Q9DTE6	Q9dte6 hepatitis c
2	1784	96.9	3010	12 Q9J3F9	Q9j3f9 hepatitis c
3	1783	96.8	3010	12 P90191	P90191 hepatitis c
4	1782	96.7	3010	12 Q9J3H7	Q9j3h7 hepatitis c
5	1781	96.7	3010	12 Q9J3H5	Q9j3h5 hepatitis c
6	1780	96.6	3010	12 Q9DTE6	Q9dte6 hepatitis c
7	1778	96.5	3008	12 Q9J3F4	Q9j3f4 hepatitis c
8	1778	96.5	3010	12 P88803	P88803 hepatitis c
9	1775	96.4	3010	12 Q68788	Q68788 hepatitis c
10	1775	96.4	3010	12 Q807P3	Q807P3 hepatitis c
11	1774	96.3	3010	12 Q81755	Q81755 hepatitis c
12	1774	96.3	2284	12 Q81817	Q81817 hepatitis c
13	1774	96.3	3010	12 P89966	P89966 hepatitis c
14	1774	96.3	3010	12 Q9J3G6	Q9j3g6 hepatitis c
15	1773	96.3	3010	12 Q99AU2	Q99au2 hepatitis c
16	1773	96.3	3010	12 Q9J3H3	Q9j3h3 hepatitis c

17	1773	96.3	3010	12 Q9QIX6	Q9qix6 hepatitis c
18	1773	96.3	3010	12 Q9QIX5	Q9qix5 hepatitis c
19	1773	96.3	3010	12 Q9DTE4	Q9dte4 hepatitis c
20	1772	96.2	3010	12 Q9DTE7	Q9dte7 hepatitis c
21	1771	96.1	3010	12 Q9J3I0	Q9j3i0 hepatitis c
22	1770	96.1	3010	12 Q9QIX3	Q9qix3 hepatitis c
23	1770	96.1	3010	12 Q09796	Q09796 hepatitis c
24	1769	96.0	3010	12 Q9QIX5	Q9qix5 hepatitis c
25	1769	96.0	3014	12 Q9DTE0	Q9dte0 hepatitis c
26	1768	96.0	3010	12 Q9DTE0	Q9dte0 hepatitis c
27	1767	95.9	3010	12 Q9J3H9	Q9j3h9 hepatitis c
28	1766	95.9	3010	12 Q9J3H0	Q9j3h0 hepatitis c
29	1766	95.9	3010	12 Q68826	Q68826 hepatitis c
30	1766	95.9	3010	12 Q9WMX2	Q9wmx2 hepatitis c
31	1765	95.8	3010	12 Q81760	Q81760 hepatitis c
32	1765	95.8	3010	12 Q9QIX8	Q9qix8 hepatitis c
33	1765	95.8	3010	12 Q9QIX7	Q9qix7 hepatitis c
34	1764	95.8	3010	12 Q02829	Q02829 hepatitis c
35	1764	95.8	3010	12 Q9J3H6	Q9j3h6 hepatitis c
36	1763	95.7	3010	12 Q9J3I1	Q9j3i1 hepatitis c
37	1761	95.6	3010	12 Q9DTE9	Q9dte9 hepatitis c
38	1761	95.6	3010	12 Q9QIX4	Q9qix4 hepatitis c
39	1761	95.6	3011	12 Q9DTE3	Q9dte3 hepatitis c
40	1759	95.5	3013	12 Q9J3H4	Q9j3h4 hepatitis c
41	1758	95.4	3010	12 Q9QIX6	Q9qix6 hepatitis c
42	1757	95.4	3010	12 Q8QRL8	Q8qrl8 hepatitis c
43	1757	95.4	3010	12 Q81541	Q81541 hepatitis c
44	1755	95.3	2864	12 Q92973	Q92973 hepatitis c
45	1755	95.3	3010	12 Q9DTE5	Q9dte5 hepatitis c

## ALIGNMENTS

RESULT 1  
ID Q9DTE6 PRELIMINARY; PRT, 3010 AA.  
AC Q9DTE6;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage, Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HCV142;  
RA Takahashi K., Iwata K., Matsumoto H., Nakao K.,  
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
RA Mishiro S.;  
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
with hepatocellular carcinoma: the 'progression score' revisited.";  
Submitted (SEP-2000) to the EMBL/Genbank/DDI databases.  
RL - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
EMBL: A8049091; BAB1804.1; -.  
CC PIR: A61196; A61196.  
CC PIR: P50329; P50329.  
CC HSSP: P26663; IUXP.  
DR GO: GO:0016021; C: integral to membrane; IEA.  
DR GO: GO:0019028; C: viral capsid; IEA.  
DR GO: GO:0019031; C: viral envelope; IEA.  
DR GO: GO:0005524; F: ATP binding; IEA.  
DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
DR GO: GO:0005489; F: electron transporter activity; IEA.  
DR GO: GO:0016787; F: hydrolase activity; IEA.  
DR GO: GO:0003723; F: RNA directed; IEA.  
DR GO: GO:0003658; F: RNA directed RNA polymerase activity; IEA.  
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR003003; Cys\_ser\_lypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV\_NSI.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NSI; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; viral\_RdRp; 1.  
 DR ProDom: PD18062; HCV\_NSI; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydroxylase; Nonstructural protein; Polypeptide; Transmembrane.  
 SC RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SC SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 97.3%; Score 1792; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 6,3e-143;  
 Matches 337; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRRAIIILTCVAHPELIPITIKLLAIGPILWLOAGITKXVYFR 60  
 DB 855 AHQVWPPLNVRGGRRAIIILTCVAHPELIPITIKLLAIGPILWLOAGITRVPYFR 914

QY 61 AAGLIRACMLVRKAGSHYQMAFMKLAALTGTYYVDHLTPLODMAHAGIRDLAVAVEV 120  
 DB 915 AAGLIRACMLVRKAGSHYQMAFMKLAALTGTYYVDHLTPLODMAHAGIRDLAVAVEV 974

QY 121 IFSDMEYKITTGADTACGDIISGLFVSARRGRREILGPADNFEQGMFLAPITAYGQ 180  
 DB 975 VFSDMEYKITTGADTACGDIISGLFVSARRGRREILGPADNFEQGMFLAPITAYGQ 1034

QY 181 QTRGLACITSLTGRKNOVEGEVQVSTATOSFATCNVGCMTVFHAGAGSKTLAAGR 240  
 DB 1035 QTRGLACITSLTGRKNOVEGEVQVSTATOSFATCNVGCMTVFHAGAGSKTLAAGR 1094

QY 241 GPITQMTNVDOQLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIIVRRGDSRGLIS 300  
 DB 1095 GPITQMTNVDOQLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIIVRRGDSRGLIS 1154

QY 301 PRPVSYIKSSGGPLCPGSHAVGIRRAACTGKAVKADFIIVEMETMR 352  
 DB 1155 PRPVSYIKSSGGPLCPGSHAVGIRRAACTGKAVKADFIIVEMETMR 1206

RESULT 2

Q9J3F9 ID Q9J3F9 PRELIMINARY; PRT; 3010 AA.  
 AC Q9J3F9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCB1:TaxID=11103;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=MD3;  
 CC Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;  
 CC "Characteristics of hepatitis C viral genome associated with disease  
 CC progression."  
 CC Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: A  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AF207774; AAF65964.1; -  
 CC PIR: A61196; A61196.  
 CC PIR: P0246; P0246.  
 CC PIR: PS0329; PS0329.  
 CC HSR: P27958; 1HE1.  
 CC MEROPS: S29.001; -  
 DR MEROPS; U39.001; -  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR00345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NSI.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NSI; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; viral\_RdRp; 1.  
 DR ProDom: PD18062; HCV\_NSI; 1.

DR SMART, SM00487, DEXDC, 1.  
 DR PROSITE, PS00190, CYTOCHROME C, 1.  
 KM Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327102 MW; 71629CDB93E6E0C7 CRC64;

Query Match 96.9%; Score 1784; DB 12; Length 3010;  
 Best Local Similarity 95.2%; Pred. No. 3e-142;  
 Matches 335; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRDAIILITCAVPELIPDITKLLAIFGLMVLQAGITKVPYFVR 60  
 DB 855 AHLHWVPLNVRGGRDAIILITCAVPELIPDITKLLAIFGLMVLQAGITKVPYFVR 914  
 QY 61 AAGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHTPLQDMAHGLRLAFAVEPV 120  
 DB 915 AAGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHTPLQDMAHGLRLAFAVEPV 974  
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVARSREIILGPADNFEQGWRLAPITAYSQ 180  
 DB 975 VFSDEMKIITWGADTAACGDIISGLPVARSREIILGPADNFEQGWRLAPITAYSQ 1034  
 QY 181 QTRGLIGCIITSLTGRDKQVEGEVQVSTAFQSFATCVNGVCMVFFHAGSKTLAEPK 240  
 DB 1035 QTRGLIGCIITSLTGRDKQVEGEVQVSTAFQSFATCVNGVCMVFFHAGSKTLAEPK 1094  
 QY 241 GPITQMTYNDQDVLVGMQAPPGARSMPTCTGSSDLYLVTRHADVTVRRRGDSRGLLS 300  
 DB 1095 GPITQMTYNDQDVLVGMQAPPGARSMPTCTGSSDLYLVTRHADVTVRRRGDSRGLLS 1154  
 QY 301 PRPVSYKSGSSGGLTSPSGHAGVIFRAAVCTRGAAKAVFIPVESHMETMR 352  
 DB 1155 PRPVSYKSGSSGGLTSPSGHAGVIFRAAVCTRGAAKAVFIPVESHMETMR 1206

RESULT 3  
 ID P90191 PRELIMINARY; PRT; 3010 AA.  
 AC P90191;  
 DT 01-MAY-1997 (TRENBLREL. 03, Created)  
 DT 01-MAY-1997 (TRENBLREL. 03, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA Enomoto N.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RA MEDLINE=95340824; PubMed=7542279;  
 RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,  
 RA Yamamoto C., Izumi N., Marumo F., Sato C.;  
 RT "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b";  
 RL J. Clin. Invest. 96:224-230 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: D50482; BAA09073.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: P00254; P00254.  
 DR PIR: P00804; P00804.  
 DR PIR: P00329; P00329.  
 DR PDB: 1DXW; 12-JAN-01.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.

DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_cysteine.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002511; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001450; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01538; HCV\_NS1; 1.  
 DR Pfam: PF01537; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KM Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 FT CHAIN 1  
 FT CHAIN 191  
 FT CHAIN 192  
 FT CHAIN 384  
 FT CHAIN 809  
 FT CHAIN 810  
 FT CHAIN 1026  
 FT CHAIN 1027  
 FT CHAIN 1657  
 FT CHAIN 1658  
 FT CHAIN 1711  
 FT CHAIN 1712  
 FT CHAIN 1973  
 FT CHAIN 2419  
 FT CHAIN 2420  
 FT CHAIN 3010  
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 96.9%; Score 1783; DB 12; Length 3010;  
 Best Local Similarity 95.5%; Pred. No. 3.7e-142;  
 Matches 336; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRDAIILITCAVPELIPDITKLLAIFGLMVLQAGITKVPYFVR 60  
 DB 855 AHLHWVPLNVRGGRDAIILITCAVPELIPDITKLLAIFGLMVLQAGITKVPYFVR 914  
 QY 61 AAGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHTPLQDMAHGLRLAFAVEPV 120  
 DB 915 AAGLIRACMLVRKAGGHVYQMAFMKLAALTGTYYVDHTPLQDMAHGLRLAFAVEPV 974  
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVARSREIILGPADNFEQGWRLAPITAYSQ 180  
 DB 975 VFSDEMKIITWGADTAACGDIISGLPVARSREIILGPADNFEQGWRLAPITAYSQ 1034  
 QY 181 QTRGLIGCIITSLTGRDKQVEGEVQVSTAFQSFATCVNGVCMVFFHAGSKTLAEPK 240  
 DB 1035 QTRGLIGCIITSLTGRDKQVEGEVQVSTAFQSFATCVNGVCMVFFHAGSKTLAEPK 1094



DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0016787; F:hydrolase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006508; P:electron transport; IEA.  
 DR GO:0006350; P:proteolysis and peptidolysis; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral genome transformation; IEA.  
 DR InterPro: IPR003003; Cys Ser tyrosin.  
 DR InterPro: IPR00345; Cyt\_c\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: PS00190; CYTOCHROME\_C; 1.  
 DR ATP-binding; Coat protein; Glycoprotein; Helicase;  
 KM Hydrolyase; Nonstructural protein; Polypeptide;  
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B93B7AA4B CRC64;

Query Match 96.7%; Score 1781; DB 12; Length 3010;  
 Best Local Similarity 95.5%; Pred. No. 5.5e-142;  
 Matches 336; Conservative 8; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 AHQVWIPPLNVRGGRDAIILLTCVAPPELIFDITKLLAIFGLMVLQGITKVPYFVR 60  
 DB 855 AHQVWIPPLNVRGGRDAIILLTCVAPPELIFDITKLLAIFGLMVLQGITRVPYFVR 914  
 QY 61 AAGIIRACMLVRKAGHYVOMAFMKALATGTVVYHPLPLDMAGLRDAVAPEV 120  
 DB 915 AAGIIRACMLVRKAGHYVOMAFMKALATGTVVYHPLPLDMAGLRDAVAPEV 974  
 QY 121 IFSDMVKIITWGAADTAACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 180  
 DB 975 VFSMEKRIITWGAADTAACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 1034  
 QY 181 QTRRLGLCITSLTGRDNQVREGVQVYVATQSFATCVNGVYVYFHAGSKTLAEPK 240  
 DB 1035 QTRRLGLCITSLTGRDNQVREGVQVYVATQSFATCVNGVYVYFHAGSKTLAEPK 1094  
 QY 241 GPIITQMTNVDDLVGVQAPPGASMTPTCGSSDLYLVTRHADVIVRRRGDSRGLLS 300  
 DB 1095 GPIITQMTNVDDLVGVQAPPGASMTPTCGSSDLYLVTRHADVIVRRRGDSRGLLS 1154

QY 301 PEPVSYLKSSGGPLCPGSHAVGIFRAVCTRGVAKAVDEIPVESMETTR 352  
 DB 1155 PEPVSYLKSSGGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETTR 1206  
 RESULT 6  
 ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.  
 AC Q9DTE6  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA, positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 CX NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV221;  
 RA Takahashi K., Iwata X., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 with hepatocellular carcinoma: the 'progression score' revisited";  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBCC databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: AB049101; BAB18814.1; --  
 DR PIR; A61196; A61196.  
 DR PIR; PQ0246; PQ0246.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; IJXP.  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0019031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0016787; F:hydrolase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006508; P:electron transport; IEA.  
 DR GO:0006350; P:proteolysis and peptidolysis; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral genome transformation; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser tyrosin.  
 DR InterPro: IPR00345; Cyt\_c\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002166; HCV\_NS5a.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF01602; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR ATP-binding; Coar protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydrolyase; Nonstructural protein; Polyprotein;  
 KM RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SO SEQUENCE 3010 AA; 327108 MW; DE182D810EF78E4 CRC64;

Query Match 96.5%; Score 1780; DB 12; Length 3010;  
 Best Local Similarity 95.7%; Pred. No. 6.6e-142;  
 Matches 337; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQNGITKVEYFVR 60  
 DB 855 AHLHWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQNGITKVEYFVR 914  
 QY 61 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLPLODMAHAGRLDAVAEVPV 120  
 DB 915 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLPLODMAHAGRLDAVAEVPV 974  
 QY 121 IFSMEKILITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 180  
 DB 975 VFSMEKILITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 1034  
 QY 181 QTRGLGCIITSLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 240  
 DB 1035 QTRGLGCIITSLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 1094  
 QY 241 GPIQMTNTVDQLVGMQAPPGASMTPTCTGSSDYLVTNRHADVIVRRRGDSRGLS 300  
 DB 1095 GPIQMTNTVDQLVGMQAPPGASMTPTCTGSSDYLVTNRHADVIVRRRGDSRGLS 1154  
 QY 301 PRPVSYLKGSSGGELCPGSHAVGIFRAVCTRGVAKAVDFIVESMETTMR 352  
 DB 1155 PRPVSYLKGSSGGELCPGSHAVGIFRAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 7  
 Q9J3F4 PRELIMINARY; PRT; 3008 AA.

AC Q9J3F4; 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CC NCBI\_Taxid=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD34;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease  
 progression";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases  
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC EMBL; AF208024; AAF61205.1; -  
 DR EMBL; AF208024; AAF61205.1; -  
 DR PIR; P00246; P00246;  
 DR PIR; P00246; P00246;  
 DR PIR; P00246; P00246;  
 DR HSP; P26653; IUXP.

DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F: electron transporter activity; IEA.  
 DR GO; GO:0016787; F: hydrolyase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008235; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005196; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006118; F: electron transport; IEA.  
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_trypsin.  
 DR InterPro; IPR000345; Cys\_ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV capsid.  
 DR InterPro; IPR002519; HCV capsid.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002688; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR001650; Helicase C.  
 DR InterPro; IPR004109; peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PS.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF02907; HCV NS4a; 1.  
 DR Pfam; PF01006; HCV NS4b; 1.  
 DR Pfam; PF01001; HCV NS5a; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; Helicase C; 1.  
 DR Pfam; PF00998; Viral RdRp; 1.  
 DR Pfam; PF01602; HCV NS1; 1.  
 DR PRODOM; PD186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR ATP-binding; Coar protein; Envelope protein; Glycoprotein; Helicase;  
 KM Hydrolyase; Nonstructural protein; Polyprotein;  
 KM RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SO SEQUENCE 3008 AA; 326834 MW; 99AE09E14C3109F4 CRC64;

Query Match 96.5%; Score 1778; DB 12; Length 3008;  
 Best Local Similarity 95.5%; Pred. No. 9.8e-142;  
 Matches 336; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQNGITKVEYFVR 60  
 DB 853 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQNGITKVEYFVR 912  
 QY 61 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLPLODMAHAGRLDAVAEVPV 120  
 DB 913 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLPLODMAHAGRLDAVAEVPV 972  
 QY 121 IFSMEKILITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 180  
 DB 973 VFSMEKILITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 1032  
 QY 181 QTRGLGCIITSLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 240  
 DB 1033 QTRGLGCIITSLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 1092  
 QY 241 GPIQMTNTVDQLVGMQAPPGASMTPTCTGSSDYLVTNRHADVIVRRRGDSRGLS 300

Db 1093 GPTITMTNVDDLVGMQAPPGARSLTPTCTCGSSDLYLTRADVIPIVRRDGRSRLS 1152  
 QY 301 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMETTMR 352  
 Db 1153 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMETTMR 1204

RESULT 8  
 P88803 PRELIMINARY; PRT; 3010 AA.  
 ID P88803  
 AC P88803; 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Genome polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=HCV-1b;  
 RA Enomoto N.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b; PubMed=7542279;  
 RA MEDLINE=95340824; Asahina Y., Kurosaki M., Murakami T.,  
 RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,  
 RA Yamamoto C., Izumi N., Marumo F., Sato C.;  
 RT "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b.";  
 RL J. Clin. Invest. 96:224-230 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 CC EMBL: D50484; BAA09075.1; -  
 DR PIR: A61196; A61196.  
 DR HSP: P26663; INS3.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.

DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02807; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD185062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; transferase; Transmembrane.  
 FT CHAIN 1 191  
 FT CHAIN 192 383  
 FT CHAIN 384 809  
 FT CHAIN 810 1026  
 FT CHAIN 1027 1657  
 FT CHAIN 1658 1711  
 FT CHAIN 1712 1972  
 FT CHAIN 1973 2419  
 FT CHAIN 2420 3010  
 SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEFB8 CRC64;

Query Match 96.5%; Score 1778; DB 12; Length 3010;  
 Best Local Similarity 94.9%; Pred. No. 9,8e-142;  
 Matches 334; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 AHQWIPPLNVRGGGDAIILITCAVHPILFDITKLLAIFGPMVLQAGITKYFVR 60  
 Db 855 AHQWIPPLNVRGGGDAIILITCAVHSELVEITKILAIPTGPMVLQAGIRPYFVR 914  
 QY 61 AGLIRACMLVKAAGHYQVAFKKAALTGTYYDHLTPLODMAHAGRLDAVAVEV 120  
 Db 915 AGLIRVCMVLRKVGAGHYQVAFVLAALTGTYYNHLLPQDMAHTGLRLDAVAVEV 974  
 QY 121 ISDMVKIITGAPTLAAGDIIISGLPVARGREILIGPANEFGQGRLLAPITAYSQ 180  
 Db 975 VSDMETKIITGADPLAACGDIISGLPVARGREILIGPANEFGQGRLLAPITAYSQ 1034  
 QY 181 QTRGLGCIITLTGDKNQVGEVQVSTATQSLATCVNGVCTVFEAGSKTLAEPK 240  
 Db 1035 QTRGLGCIITLTGDKNQVGEVQVSTATQSLATCVNGVCTVYHGAASKTLAEPK 1094  
 QY 241 GPTITMTNVDDLVGMQAPPGARSLTPTCTCGSSDLYLTRADVIPIVRRDGRSRLS 300  
 Db 1095 GPTITMTNVDDLVGMQAPPGARSLTPTCTCGSSDLYLTRADVIPIVRRDGRSRLS 1154

Db 301 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMETTMR 352  
 Db 1153 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 9  
 O68788 PRELIMINARY; PRT; 3010 AA.  
 ID O68788  
 AC O68788;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE HCV polyprotein (Genome polypeptide).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP MEDLINE=96362158; PubMed=8720135;  
 RA Seki M., Honda Y.;  
 RT "Phosphorothioate antisense oligodeoxynucleotides capable of  
 RT inhibiting Hepatitis C virus gene expression: In vitro translation  
 RT assay.";  
 RL J. Biochem. 118:1199-1204 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: D45172; BAA08120.1; -  
 DR PIR: A61196; A61196  
 DR PIR: P00246; P00246  
 DR PIR: P0329; P0329  
 DR HSSP: P26663; 1UXP.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008265; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F: transcription; IEA.  
 DR GO: GO:0019079; F: viral genome replication; IEA.  
 DR GO: GO:0019087; F: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_lysine.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV capsid.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00398; Viral\_RdRp; 1.  
 DR Pfam: PF01866; HCV\_NS1; 1.  
 DR Pfam: PF01866; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SO SEQUENCE 3010 AA; 326880 MW; EED840E6A050E766 CCG64;

Query Match 96.4%; Score 1775; DB 12; Length 3010;  
 Best Local Similarity 94.9%; Pred. No. 1.8e-141;  
 Matches 334; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGLQWVPPINVRGGRDAIILITCAVPELIPITKLLAI FGLPLWLQGIKVPYFVR 60  
 DB 855 AHIQWVPPINVRGGRDAIILITCAVPELIPITKLLAI FGLPLWLQGIKVPYFVR 914  
 QY 61 AGLIRACMLVRAAGHYVQMAFMKLAALITGVYDHLTPLODMAHAGIRDLAAVEPV 120  
 DB 915 AGLIRACMLVRAAGHYVQMAFMKLAALITGVYDHLTPLODMAHAGIRDLAAVEPV 974  
 QY 121 IFSMEVKIITWGDITACGDIISGLVSARSGEILGPDNFEQGMRLAPITAYSO 180  
 DB 975 VFSMEVKIITWGDITACGDIISGLVSARSGEILGPDNFEQGMRLAPITAYSO 1034  
 QY 181 QTRGLGCIITSLTGRDXNVEGVCVVSATOSFLATCVNGVCMVTFHAGSKTLAAGK 240  
 DB 1035 QTRGLGCIITSLTGRDXNVEGVCVVSATOSFLATCVNGVCMVTFHAGSKTLAAGK 1094

QY 241 GPITQMTNVDQDLVGMGAPPARGASMTCTCGSSDLVYVTRHADVIPVRRGDSRGSLLS 300  
 DB 1095 GPITQMTNVDQDLVGMGAPPARGASMTCTCGSSDLVYVTRHADVIPVRRGDSRGSLLS 1154  
 QY 301 PRPVSYLKGSSGGLPCSGHVGIFRAAVCTRGVAKVDFIPVSEMTTMR 352  
 DB 1155 PRPVSYLKGSSGGLPCSGHVGIFRAAVCTRGVAKVDFIPVSEMTTMR 1206

RESULT 10  
 ID 0807P3 PRELIMINARY; PRT; 3010 AA.  
 AC 0807P3  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepcivirine.  
 OX NCBI\_TaxID=11103;  
 RN NCBI\_TaxID=11103;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MILE;  
 RX MEDLINE=22047193; PubMed=12051758;  
 RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,  
 RA Nio Y., Hosaka M., Miyamori Y., Shimotohno K.;  
 RT "Subgenomic replicon derived from a cell line infected with the  
 RT hepatitis C virus."  
 RL Biochem. Biophys. Res. Commun. 293:993-999(2002).  
 DR EMBL: AB080299; BACS4896.1; -  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0006118; F: electron transport; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F: transcription; IEA.  
 DR GO: GO:0019079; F: viral genome replication; IEA.  
 DR GO: GO:0019087; F: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_lysine.  
 DR InterPro: IPR000345; Cys\_Ser\_lysine.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV capsid.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00398; Viral\_RdRp; 1.



DR PRODOM; ED186062; HCV NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KM Polypeptide.  
 SQ SEQUENCE 3010 AA; 327097 MW; EE6418C7A7235686 CRC64;  
 Query Match 96.4%; Score 1775; DB 12; Length 3010;  
 Best Local Similarity 95.2%; Pred. No. 1.8e-141;  
 Matches 335; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRDAIILITCAVHPELFDITKLLAIIFGPLYLQAGITKVPYFVR 60  
 DB 855 AHLQWVWPLNVRGGRDAIILITCAVHPELFDITKLLAIIFGPLYLQAGITKVPYFVR 914  
 QY 61 AAGLIRACMLVNRKAGGHVQMAFMKLAALITGVVYDHLPLQDMAHAGRLDAVAVEPV 120  
 DB 915 AAGLIRACMLVNRKAGGHVQMAFMKLAALITGVVYDHLPLQDMAHAGRLDAVAVEPV 974  
 QY 121 IFSDMVEKIIITWGAADTAACGDIISGLPVSARGREILLGPADNFEQGWELAPITAYSQ 180  
 DB 975 VFSMEKRIITWGAADTAACGDIISGLPVSARGREILLGPADNFEQGWELAPITAYSQ 1034  
 QY 181 QTRBLGCIITSLTGRDNQVGEVQVYVSTATOSFLATCNVGVWTVFHGAGSKTLAIPK 240  
 DB 1035 QTRBLGCIITSLTGRDNQVGEVQVYVSTATOSFLATCNVGVWTVFHGAGSKTLAIPK 1094  
 QY 241 GPIQMTNVDQDVLVGMQAPGASMTPTCGSSDLYLVTRHADVIFVRSGSGSLLS 300  
 DB 1095 GPIQMTNVDQDVLVGMQAPGASMTPTCGSSDLYLVTRHADVIFVRSGSGSLLS 1154  
 QY 301 PRPVSYLKSGSGGELLCPSGHAGVIFRAVCTRGVANAVIDFIPESMTTMR 352  
 DB 1155 PRPVSYLKSGSGGELLCPSGHAGVIFRAVCTRGVANAVIDFIPESMTTMR 1206

RESULT 11  
 ID 081755 PRELIMINARY; ERT; 1186 AA.  
 AC 081755;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Polypeptide (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91088550; PubMed=2175503;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RA "Molecular cloning of the human hepatitis C virus genome from Japanese  
 RT patients with non-A, non-B hepatitis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9204440; PubMed=1658196;  
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,  
 RA Miyakawa Y., Mayumi M.;  
 RA "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91140698; PubMed=1847440;  
 RA Takamizawa A., Mori C., Kanabe S., Murakami S., Fujita J., Onishi E.,  
 RA Andoh T., Yoshida I., Okayama H.;  
 RA "The structure and organization of the Hepatitis C virus genome  
 RT isolated from human carriers.";  
 RL J. Virol. 65:1105-1113(1991).

RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos C., Coit D., Medina-Seiby A., Barr P.J., Weiner A.,  
 RA Bradley D.W., Kuo G., Houghton M.;  
 RA "Genetic organization and diversity of the hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230206; PubMed=1314445;  
 RA Chen P., Lin M., Tai K., Lin P., Lin C., Chen D.;  
 RT "The Taiwanese hepatitis C virus genome: Sequence determination and  
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113(1992).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230232; PubMed=1314459;  
 RA Okamoto H., Kurai K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: Comparative study of four distinct  
 RT genotypes.";  
 RL Virology 188:331-341(1992).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93323208; PubMed=8392606;  
 RA Hijikata M., Mizushima H., Akagi T., Mori S., Kakuchi N., Kato N.,  
 RA Tanaka T., Kimura K., Shimotohno K.;  
 RT "Two distinct proteinase activities required for the processing of a  
 RT putative nonstructural precursor protein of hepatitis C virus.";  
 RL J. Virol. 67:4665-4675(1993).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Hijikata M.;  
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D11397; BAA20975.1; -.  
 DR PIR; A61186; A61186.  
 DR PIR; P80329; P80329.  
 DR PDB; 1DXP; 28-MAR-02.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; F:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR001650; Helicase\_C29.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW ATP-binding; Helicase; Hydrolase.  
 FT NON TER  
 SQ SEQUENCE 1186 AA; 126280 MW; 34170478BA23729A CRC64;

Query Match 96.3%; Score 1774; DB 12; Length 1186;  
 Best Local Similarity 95.2%; Pred. No. 6.4e-142;  
 Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRDAIILITCAVHPELFDITKLLAIIFGPLYLQAGITKVPYFVR 60  
 DB 133 AHLQWVWPLNVRGGRDAIILITCAVHPELFDITKLLAIIFGPLYLQAGITKVPYFVR 192  
 QY 61 AAGLIRACMLVNRKAGGHVQMAFMKLAALITGVVYDHLPLQDMAHAGRLDAVAVEPV 120

```

Db 193 AAGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLDMAHAGRDIAVAVEPV 252
QY 121 IFSDEVKIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180
Db 253 VFSDEMETKIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 312
QY 181 QTRGLGCIITSLTGRKNOVEGVYVSTATOSFLATCNGVCMVTFHAGSKTLTAGPK 240
Db 313 QTRGLGCIITSLTGRKNOVEGVYVSTATOSFLATCNGVCMVTFHAGSKTLTAGPK 372
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLLS 300
Db 373 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLLS 432
QY 301 PRPVSYLKSSGGGGLLCPSGHVAVGFPAVCTRGVAKAVDFIVESMETTMR 352
Db 433 PRPVSYLKSSGGGGLLCPSGHVAVGFPAVCTRGVAKAVDFIVESMETTMR 484

RESULT 12
081817 PRELIMINARY; PRT; 2284 AA.
ID 081817
AC 081817
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polypeptide precursor (Genome polypeptide).
OS Hepatitis C virus type 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_TaxID=40271;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=9406848; PubMed=7504283;
RA Hijikata M., Mitsuhashi H., Tanji Y., Komada Y., Hirowatari Y.,
RA Akagi T., Kimura K., Shimotohno K.;
RT "Proteolytic processing and membrane association of putative
RT nonstructural proteins of hepatitis C virus."
RL Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=9433810; PubMed=8056334;
RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
RT "Identification of the domain required for trans-cleavage activity of
RT hepatitis C viral serine proteinase."
RL Gene 145:215-219(1994).
RN 13
RP SEQUENCE FROM N.A.
RX MEDLINE=9505678; PubMed=7966638;
RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
RT "Hepatitis C virus polypeptide processing: kinetics and mutagenic
RT analysis of serine proteinase-dependent cleavage."
RL J. Virol. 68:8418-8422(1994).
RN 14
RP SEQUENCE FROM N.A.
RX MEDLINE=9515683; PubMed=7853491;
RA Tanji Y., Hijikata M., Sato S., Kaneko T., Shimotohno K.;
RT "Hepatitis C virus-encoded nonstructural protein NS4A has versatile
RT functions in viral protein processing."
RL J. Virol. 69:1575-1581(1995).
DR EMBL; D16435; BAA03905.1; -
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00246; P00246.
DR PIR; P00246; P00246.
DR HSBP; P26663; IXP.
DR GO; GO:0019012; C:viral; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003688; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008266; F:serine-type peptidase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

```

```

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_lysin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001450; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVlr.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR SMART; SM00487; DEXDC; 1.
KM Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;
KW Signal; transferase.
FT SIGNAL 20
FT SIGNAL 4
FT SIGNAL 55
FT CHAIN 21
FT CHAIN 84
FT CHAIN 301
FT CHAIN 932
FT CHAIN 986
FT CHAIN 1247
FT CHAIN 1694
FT CHAIN 2284
SQ SEQUENCE 2284 AA; 247221 MW; DC272A1517046337 CRC64;

Query Match
Best Local Similarity 96.3%; Score 1774; DB 12; Length 2284;
Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHLQWVPPVPLNVRGRDAIILLTCVHBEILFDITKLLALFGLMVLQAGITKVPYFVR 60
Db 129 AHLQWVPPVPLNVRGRDAIILLTCVHBEILFDITKLLALFGLMVLQAGITKVPYFVR 188
QY 61 AAGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLDMAHAGRDIAVAVEPV 120
Db 189 AAGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLDMAHAGRDIAVAVEPV 248
QY 121 IFSDEVKIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180
Db 249 VFSDEMETKIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 308
QY 181 QTRGLGCIITSLTGRKNOVEGVYVSTATOSFLATCNGVCMVTFHAGSKTLTAGPK 240
Db 309 QTRGLGCIITSLTGRKNOVEGVYVSTATOSFLATCNGVCMVTFHAGSKTLTAGPK 368
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLLS 300
Db 369 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLLS 428
QY 301 PRPVSYLKSSGGGGLLCPSGHVAVGFPAVCTRGVAKAVDFIVESMETTMR 352
Db 429 PRPVSYLKSSGGGGLLCPSGHVAVGFPAVCTRGVAKAVDFIVESMETTMR 480

RESULT 13
P89966 PRELIMINARY; PRT; 3010 AA.
ID P89966
AC P89966
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RNA for polypeptide (Genome polypeptide).

```

OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CC NCBI\_Taxid=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=type 1b;  
 RA Tanaka T.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=type 1b;  
 RA TANAKA T.;  
 RT "TMORF.";  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL; D89872; BAA14035.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00804; P00804.  
 DR PIR; P50329; P50329.  
 DR HSP; P26663; IUXP.  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008226; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV env.  
 DR InterPro; IPR002519; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR002868; HCV NS5a.  
 DR InterPro; IPR002166; HCV RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4; 1.  
 DR Pfam; PF01001; HCV NS4b; 1.  
 DR Pfam; PF01506; HCV NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; viral\_RdRp; 1.  
 DR ProDom; PD186062; HCV\_NSL; 1.  
 DR SMART; SM00467; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 337023 MW; E075BD9CFD8D1261 CRC64;

Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 AHLOWIPPLNVRGGRRDAITLLTCVHPPELLIDIKLLATFGLPMVQAQGTKYFVR 60  
 DB 855 AHLOWVPPPLNVRGGRRDAITLLTCVHPPELLIDIKLLATFGLPMVQAQGTKYFVR 914  
 QY 61 AAGLIRACMLVKAAGHYVQWAFMKLAALTGTYVDHLLPLQDMAHGLRDLAAVEPV 120  
 DB 915 AAGLIRACMLVKAAGHYVQWAFMKLAALTGTYVDHLLPLQDMAHGLRDLAAVEPV 974  
 QY 121 IFSDMEVKITTGADTAACGDIISGLPVSARGREIILGPANFEGCGRLAPITAYSQ 180  
 DB 975 VFSDMEKITTGADTAACGDIISGLPVSARGREIILGPANFEGCGRLAPITAYSQ 1034  
 QY 181 QTRGLGCIITSLTRDKNQVEGVVSTATQSLATCVNGVCTVPHGAGSKTLAEPK 240  
 DB 1035 QTRGLGCIITSLTRDKNQVEGVVSTATQSLATCVNGVCTVPHGAGSKTLAEPK 1094  
 QY 241 GITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVTPRRRDSRGLIS 300  
 DB 1095 GITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVTPRRRDSRGLIS 1154  
 QY 301 PRPVSTLKSGSGGPLCPGSHAVGIFRAVCTRGVAKAVDFPVSMTETKR 352  
 DB 1155 PRPVSTLKSGSGGPLCPGSHAVGIFRAVCTRGVAKAVDFPVSMTETKR 1206  
 RESULT 14  
 ID Q9J3G6 PRELIMINARY; PRT: 3010 AA.  
 AC Q9J3G6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CC NCBI\_Taxid=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD26;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease progression."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL; AF207767; AAF65957.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00254; P00254.  
 DR PIR; P50329; P50329.  
 DR HSP; P26663; IUXP.  
 DR MEROPS; S29.002; -  
 DR GO; GO:0016021; C: integral to membrane; IEA.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0019031; C: viral envelope; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F: electron transporter activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0016740; F: transferase activity; IEA.  
 DR GO; GO:0006118; P: electron transport; IEA.  
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P: transcription; IEA.  
 DR GO; GO:0019079; P: viral genome replication; IEA.  
 DR GO; GO:0019087; P: viral transformation; IEA.

DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; Peptidase C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase C; 1.  
 DR Pfam: PF00999; Viral RdRp; 1.  
 DR ProDom: PD166062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327165 MW; 74PAB6B80F24837B CRC64;

Query Match 96.3%; Score 1774; DB 12; Length 3010;  
 Best Local Similarity 95.5%; Pred. No. 2.1e-141;  
 Matches 336; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 AMLVWIPINVRGGRDAIILITCAVPELIPITIKLLAIFPPLVYLQAGITKVEYVR 60  
 Db 855 AMLQWIPINVRGGRDAIILITCAVPELIPITIKLLAIFPPLVYLQAGITKVEYVR 914

QY 61 AAGIIRACMLVRKAGAGHYOMAFMKLAALTGYYVDHLLPLQDMAHAGRDIAVVEPY 120  
 Db 915 AAGIIRACMLVRKAGAGHYOMAFMKLAALTGYYVDHLLPLQDMAHAGRDIAVVEPY 974

QY 121 IFSDMEVKIITWGDITACGDIISGLPVSARGREIILGPADNFEQGWELLAPITAYSQ 180  
 Db 975 VFSMEVKIITWGDITACGDIISGLPVSARGREIILGPADNFEQGWELLAPITAYSQ 1034

QY 181 QTRRLGCIITSLTGRKNOVEGEVYVSTATQSFATCNVGCWYVFHAGSKTLAGRK 240  
 Db 1035 QTRRLGCIITSLTGRKNOVEGEVYVSTATQSFATCNVGCWYVFHAGSKTLAGRK 1094

QY 241 GPITQWNTVNDODLVGQAPFGARSMPTCGSSDLYLVRHADVI PVRRGDSRGLLS 300  
 Db 1095 GPITQWNTVNDODLVGQAPFGARSMPTCGSSDLYLVRHADVI PVRRGDSRGLLS 1154

QY 301 PRPVSYIKSGSGGFLCPSSHAVGIFPAAYCTRGAKAVDFIPVESNETTMR 352  
 Db 1155 PRPVSYIKSGSGGFLCPSSHAVGIFPAAYCTRGAKAVDFIPVESNETTMR 1206

RESULT 15  
 Q99AU2 PRELIMINARY; PRT; 3010 AA.  
 AC Q99AU2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus type 1b.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepatitis.  
 OX NCBI\_TaxID=31647;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=chimera of HCV-BK;  
 RA Thomson M., Nascimbent M., Gonzales S., Murthy K., Rehmann B.,  
 RA Liang J.;  
 RT "Analyses of viral sequences and virus-specific immune responses  
 during serial passage of an infectious hepatitis C virus serotype 1b  
 clone in chimpanzees."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: AF333324; AAK08509.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P02246; P02246.  
 DR PIR: P08004; P08004.  
 DR PIR: P03299; P03299.  
 DR HSSP: P26663; INB3.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0005489; F: electron transporter activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006118; P: electron transport; IEA.  
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR004109; Peptidase C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00999; Viral RdRp; 1.  
 DR ProDom: PD166062; HCV NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; 1.  
 KW Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327007 MW; 053B9A653B0AB35 CRC64;

Query Match 96.3%; Score 1773; DB 12; Length 3010;  
 Best Local Similarity 94.6%; Pred. No. 2.6e-141;  
 Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

```

QY 1 AHLOWMIPPLNVRGGDDAIIILITCAVPELIFDITTKLLAIFGRLMVLQNGITKVPYFVR 60
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 855 AHLQWVWPEPLNVRGGDDAIIILITCAVPELIFDITTKLLAIIIGPLMVLQAGITRVPYFVR 914
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 AOGIIRACMLVYRKAGGHHVQMAFMKLAALTGTVVYDHLTFLQDMAHAGLRDLAVAVEPV 120
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 915 AOGIIRACMLVYRKAGGHHVQMAFMKLGALTGTIYIYHILTLPRDMAHAGLRDLAVAVEPV 974
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 IFSDMEVKIITWGADTAACGDIISGLPV SARGREIILGPADNFBGQWRLAPITAYSQ 180
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 975 VFSDEMETKIITWGADTAACGDIILGLPV SARGREIILGPADNFBGQWRLAPITAYSQ 1034
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 QTRGLLCIITSLTGRDKNOVEGEVOVSTATOSFLATCVNGVMTVYFHAGSKTLAGPK 240
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1035 QTRGLLCIITSLTGRDKNOVEGEVOVSTATOSFLATCVNGVMTVYFHAGSKTLAGPK 1094
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 GPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVI PVRRRGDSRGSLLS 300
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1095 GPITQMTYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLTRHADVI PVRRRGDSRGSLLS 1154
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 PRPVSYLKSSGGPILCPSGHANGIFRAAVCTRGVAKAVDFIYESMETTMR 352
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1155 PRPVSYLKSSGGPILCPSGHANGIFRAAVCTRGVAKAVDFIYESMETTMR 1206
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: May 6, 2004, 09:35:48  
 Job time : 33.9219 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.8154 Seconds  
(without alignments)

1315.364 Million cell updates/sec

Title: US-10-650-585-13  
Perfect score: 1842  
Sequence: 1 AHQVWIPPLNVRGGRDAI.....RGVAKAVDFPVESMETTMR 352

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgnt\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgnt\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgnt\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgnt\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgnt\_6/ptodata/2/1aa/6C.COMB.pep.\*  
6: /cgnt\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1766	95.9	2201	4	US-09-539-601-6
2	1766	95.9	2201	4	US-09-539-601-15
3	1766	95.9	3010	4	US-09-539-601-3
4	1766	95.9	3010	4	US-09-539-601-21
5	1766	95.9	3010	4	US-09-539-601-27
6	1759	95.5	3010	4	US-09-539-601-33
7	1757	95.4	1692	3	US-09-263-933-4
8	1757	95.4	1692	3	US-09-919-901-4
9	1757	95.4	2307	3	US-09-263-933-2
10	1757	95.4	2307	3	US-09-919-901-2
11	1754	95.2	1692	3	US-09-263-933-11
12	1754	95.2	1692	3	US-09-919-901-11
13	1754	95.2	2307	3	US-09-263-933-9
14	1754	95.2	2307	3	US-09-919-901-9
15	1748	94.9	3010	3	US-09-014-416-3
16	1745	94.7	1692	3	US-09-263-933-18
17	1745	94.7	1692	3	US-09-919-901-18
18	1745	94.7	2307	3	US-09-263-933-16
19	1745	94.7	2307	3	US-09-919-901-16
20	1699	92.2	2013	1	US-08-324-977-12
21	1699	92.2	2013	2	US-08-384-616-12
22	1699	92.2	2013	2	US-08-904-686A-12
23	1699	92.2	2013	3	US-09-315-850-12
24	1699	92.2	2201	4	US-08-952-981A-2
25	1699	92.2	2620	1	US-08-324-977-32
26	1699	92.2	2620	1	US-08-384-616-32
27	1699	92.2	2620	2	US-08-904-686A-32

28	1699	92.2	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1699	92.2	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1699	92.2	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1699	92.2	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1699	92.2	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1699	92.2	3010	1	US-08-324-977-2	Sequence 2, Appl
34	1699	92.2	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1699	92.2	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1699	92.2	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1699	92.2	3010	2	US-08-904-686A-2	Sequence 2, Appl
38	1699	92.2	3010	3	US-08-904-686A-14	Sequence 14, Appl
39	1699	92.2	3010	3	US-09-315-850-2	Sequence 2, Appl
40	1699	92.2	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1620	87.9	1648	1	US-08-188-281B-12	Sequence 12, Appl
42	1620	87.9	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1620	87.9	1648	5	PCT-US95-01087-12	Sequence 12, Appl
44	1620	87.9	3011	1	US-08-188-281B-1	Sequence 1, Appl
45	1620	87.9	3011	1	US-08-453-552-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-539-601-6  
; Sequence 6, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-6

Query Match 95.9%; Score 1766; DB 4; Length 2201;  
Best Local Similarity 94.3%; Pred. No. 1.2e-170;  
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY	1	AHQVWIPPLNVRGGRDAIILITCAVHPPELLIDITKLIAIFGPLVTLQAGITKVPYFVR	60
DB	46	AHQVWIPPLNVRGGRDAIILITCAVHPPELLITKILAIIGPLVTLQAGITKVPYFVR	105
QY	61	AGQLIRACVLRKAGAGHYQVAFMKLALITGYVDHLTPVQDMNAHCLBLAVAVEV	120
DB	106	AHLIRACVLRKAGAGHYQVAFMKLALITGYVDHLTPVQDMNAHCLBLAVAVEV	165
QY	121	IFSDMEVKTITGADPAACGDIISGLPVARSRGREZILIPANFEGQRLAPITAYSQ	180
DB	166	VPSDMTKVITGADPAACGDIISGLPVARSRGREZILIPANFEGQRLAPITAYSQ	225
QY	181	QTRGLIGCIITSLTGDRKQVEGEVQVSTAFQSLFATCVNGVCTVFGAGSKTLAGPX	240
DB	226	QTRGLIGCIITSLTGDRKQVEGEVQVSTAFQSLFATCVNGVCTVFGAGSKTLAGPX	285
QY	241	GPITQWYTVVDIVGMAQPGARSMTPCTGSSSDILVTRADYVTVRRRSDSGSLIS	300
DB	286	GPITQWYTVVDIVGMAQPGARSMTPCTGSSSDILVTRADYVTVRRRSDSGSLIS	345
QY	301	PRPVSYLKSSGGPLICPSGAHVGFRAVACTRGVAKAVDFPVESMETTMR	352
DB	346	PRPVSYLKSSGGPLICPSGAHVGFRAVACTRGVAKAVDFPVESMETTMR	397

RESULT 2

US-09-539-601-15  
 ; Sequence 15, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlagel, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 15  
 ; LENGTH: 2201  
 ; TYPE: PRF  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-15

Query Match 95.9%; Score 1766; DB 4; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-170;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHTQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 60  
 DB 46 AHTQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 105  
 QY 61 AQLIRACMLVRKAGGHHYQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAVAEVPY 120  
 DB 106 AHGLIRACMLVRKAGGHHYQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAVAEVPY 165  
 QY 121 IFSMEVKIITWGDATACGDIISGLPVSARGREIILGPADNPEGQWRLLAITYASQ 180  
 DB 166 VFSDEMETKITWGDATACGDIISGLPVSARGREIILGPADNPEGQWRLLAITYASQ 225  
 QY 181 QTRGLGCIITSLGRDNQVEGVYSTATOSFLATCNVGVCMVTFHAGSKTLGPK 240  
 DB 226 QTRGLGCIITSLGRDNQVEGVYSTATOSFLATCNVGVCMVTFHAGSKTLGPK 285  
 QY 241 GPTQMTNNVDQDVGWQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLLS 300  
 DB 286 GPTQMTNNVDQDVGWQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLLS 345  
 QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 352  
 DB 346 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 397

RESULT 3  
 US-09-539-601-3  
 ; Sequence 3, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlagel, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 3010  
 ; TYPE: PRF  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-3

Query Match 95.9%; Score 1766; DB 4; Length 3010;  
 Best Local Similarity 94.3%; Pred. No. 1.9e-170;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHTQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 60

DB 855 AHTQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 914  
 QY 61 AQLIRACMLVRKAGGHHYQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAVAEVPY 120  
 DB 915 AHGLIRACMLVRKAGGHHYQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAVAEVPY 974  
 QY 121 IFSMEVKIITWGDATACGDIISGLPVSARGREIILGPADNPEGQWRLLAITYASQ 180  
 DB 975 VFSDEMETKITWGDATACGDIISGLPVSARGREIILGPADNPEGQWRLLAITYASQ 1034  
 QY 181 QTRGLGCIITSLGRDNQVEGVYSTATOSFLATCNVGVCMVTFHAGSKTLGPK 240  
 DB 1035 QTRGLGCIITSLGRDNQVEGVYSTATOSFLATCNVGVCMVTFHAGSKTLGPK 1094  
 QY 241 GPTQMTNNVDQDVGWQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLLS 300  
 DB 1095 GPTQMTNNVDQDVGWQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLLS 1154  
 QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 352  
 DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 1206

RESULT 4  
 US-09-539-601-21  
 ; Sequence 21, Application US/09539601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartschlagel, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 21  
 ; LENGTH: 3010  
 ; TYPE: PRF  
 ; ORGANISM: Hepatitis C virus  
 US-09-539-601-21

Query Match 95.9%; Score 1766; DB 4; Length 3010;  
 Best Local Similarity 94.3%; Pred. No. 1.9e-170;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHTQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 60  
 DB 855 AHTQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 914  
 QY 61 AQLIRACMLVRKAGGHHYQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAVAEVPY 120  
 DB 915 AHGLIRACMLVRKAGGHHYQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAVAEVPY 974  
 QY 121 IFSMEVKIITWGDATACGDIISGLPVSARGREIILGPADNPEGQWRLLAITYASQ 180  
 DB 975 VFSDEMETKITWGDATACGDIISGLPVSARGREIILGPADNPEGQWRLLAITYASQ 1034  
 QY 181 QTRGLGCIITSLGRDNQVEGVYSTATOSFLATCNVGVCMVTFHAGSKTLGPK 240  
 DB 1035 QTRGLGCIITSLGRDNQVEGVYSTATOSFLATCNVGVCMVTFHAGSKTLGPK 1094  
 QY 241 GPTQMTNNVDQDVGWQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLLS 300  
 DB 1095 GPTQMTNNVDQDVGWQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRGRSGSLLS 1154

QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 352  
 DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 1206



RESULT 5  
 US-09-539-601-27  
 : Sequence 27, Application US/09539601C  
 : Patent No. 6630343  
 : GENERAL INFORMATION:  
 : APPLICANT: Bartschlagel, Ralf FM  
 : TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 : FILE REFERENCE: all sequences  
 : CURRENT APPLICATION NUMBER: US/09/539,601C  
 : EARLIER FILING DATE: 2001-08-30  
 : EARLIER FILING DATE: 1991-15-178.4 GERMANY  
 : NUMBER OF SEQ ID NOS: 51  
 : SOFTWARE: Patentin Ver. 2.1  
 : SEQ ID NO 27  
 : LENGTH: 3010  
 : TYPE: PRT  
 : ORGANISM: Hepatitis C virus  
 US-09-539-601-27

Query Match 95.9%; Score 1766; DB 4; Length 3010;  
 Best Local Similarity 94.3%; Pred. No. 1.9e-170;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLCVWIPPLNVRGGRDAIILLTCVHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 60  
 DB 855 AHLCVWIPPLNVRGGRDAIILLTCVHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 914  
 QY 61 AGLIACMLVRKAGGHVQVMAFMKLAALTGYVYDHTPLQDMHAGRLDAVAEVPV 120  
 DB 915 AGLIACMLVRKAGGHVQVMAFMKLAALTGYVYDHTPLQDMHAGRLDAVAEVPV 974  
 QY 121 IFSDEVKIITWGADTAACGDIISGLPVASRGREIILGPADNFGQWRLLAPITAYSQ 180  
 DB 975 VFSDEMKIITWGADTAACGDIISGLPVASRGREIILGPADNFGQWRLLAPITAYSQ 1034  
 QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCTVPHAGSKTLAGPK 240  
 DB 1035 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCTVPHAGSKTLAGPK 1094  
 QY 241 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIPIVRRGDSRGLLS 300  
 DB 1095 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIPIVRRGDSRGLLS 1154  
 QY 301 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 352  
 DB 1155 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 6  
 US-09-539-601-33  
 : Sequence 33, Application US/09539601C  
 : Patent No. 6630343  
 : GENERAL INFORMATION:  
 : APPLICANT: Bartschlagel, Ralf FM  
 : TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 : FILE REFERENCE: all sequences  
 : CURRENT APPLICATION NUMBER: US/09/539,601C  
 : EARLIER FILING DATE: 2001-08-30  
 : EARLIER FILING DATE: 1991-15-178.4 GERMANY  
 : NUMBER OF SEQ ID NOS: 51  
 : SOFTWARE: Patentin Ver. 2.1  
 : SEQ ID NO 33  
 : LENGTH: 3010  
 : TYPE: PRT  
 : ORGANISM: Hepatitis C virus  
 US-09-539-601-33

Query Match 95.5%; Score 1759; DB 4; Length 3010;  
 Best Local Similarity 94.0%; Pred. No. 9.8e-170;  
 Matches 331; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLCVWIPPLNVRGGRDAIILLTCVHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 60  
 DB 855 AHLCVWIPPLNVRGGRDAIILLTCVHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 914  
 QY 61 AGLIACMLVRKAGGHVQVMAFMKLAALTGYVYDHTPLQDMHAGRLDAVAEVPV 120  
 DB 915 AGLIACMLVRKAGGHVQVMAFMKLAALTGYVYDHTPLQDMHAGRLDAVAEVPV 974  
 QY 121 IFSDEVKIITWGADTAACGDIISGLPVASRGREIILGPADNFGQWRLLAPITAYSQ 180  
 DB 975 VFSDEMKIITWGADTAACGDIISGLPVASRGREIILGPADNFGQWRLLAPITAYSQ 1034  
 QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCTVPHAGSKTLAGPK 240  
 DB 1035 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCTVPHAGSKTLAGPK 1094  
 QY 241 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIPIVRRGDSRGLLS 300  
 DB 1095 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIPIVRRGDSRGLLS 1154  
 QY 301 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 352  
 DB 1155 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 7  
 US-09-263-933-4  
 : Sequence 4, Application US/09263933  
 : Patent No. 6280940  
 : GENERAL INFORMATION:  
 : APPLICANT: Potts, Karen E.  
 : APPLICANT: Jackson, Roberta L.  
 : TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 : FILE REFERENCE: 0125-0005A  
 : CURRENT APPLICATION NUMBER: US/09/263,933  
 : EARLIER FILING DATE: 1999-03-08  
 : EARLIER FILING DATE: 09/129,611  
 : NUMBER OF SEQ ID NOS: 33  
 : SOFTWARE: Patentin Ver. 2.0  
 : SEQ ID NO 4  
 : LENGTH: 1692  
 : TYPE: PRT  
 : ORGANISM: Artificial Sequence  
 US-09-263-933-4

Query Match 95.4%; Score 1757; DB 3; Length 1692;  
 Best Local Similarity 93.8%; Pred. No. 6.7e-170;  
 Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLCVWIPPLNVRGGRDAIILLTCVHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 60  
 DB 134 AHLCVWIPPLNVRGGRDAIILLTCVHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 193  
 QY 61 AGLIACMLVRKAGGHVQVMAFMKLAALTGYVYDHTPLQDMHAGRLDAVAEVPV 120  
 DB 194 AGLIACMLVRKAGGHVQVMAFMKLAALTGYVYDHTPLQDMHAGRLDAVAEVPV 253  
 QY 121 IFSDEVKIITWGADTAACGDIISGLPVASRGREIILGPADNFGQWRLLAPITAYSQ 180  
 DB 254 VFSDEMKIITWGADTAACGDIISGLPVASRGREIILGPADNFGQWRLLAPITAYSQ 313  
 QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCTVPHAGSKTLAGPK 240  
 DB 314 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCTVPHAGSKTLAGPK 373  
 QY 241 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIPIVRRGDSRGLLS 300  
 DB 374 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIPIVRRGDSRGLLS 433  
 QY 301 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFPVESMETTMR 352

Db 434 PRPVSYLKSSGGPILCPSSGHAAGIFRAAVCTRGVAADVPFVESHMETMR 485

## RESULT 8

US-09-919-901-4  
Sequence 4, Application US/09919901

Patent No. 6599738  
GENERAL INFORMATION:

APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919, 901

CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263, 933

PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129, 611

PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4

LENGTH: 1692  
TYPE: PRT

ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: :  
US-09-919-901-4

Query Match 95.4%; Score 1757; DB 4; Length 1692;  
Best Local Similarity 93.8%; Pred. No. 6, 7e-170;

Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILTCVHPELFDITKLLAIFGLMVLQAGITKVPYFVR 60  
DB 134 AHLHWIPLNARGRDAIILMCAVHPELFDITKLLAIFGLMVLQAGITKVPYFVR 193  
QY 61 AAGLIRACMLVRKAAAGHYVQMAFMKLAALGTYYDHLTFLQDMANAGRDIAVAEPV 120  
DB 194 AAGLIHACMLVRKAAAGHYVQMAFMKLAALGTYYDHLTFLQDMANAGRDIAVAEPV 253  
QY 121 IFSDEVKIITWGDADTAACGDIISGLPVSARKEIILGPADNFEQGMRLLAPITAYSQ 180  
DB 254 VFSDEVKIITWGDADTAACGDIISGLPVSARKEIILGPADNFEQGMRLLAPITAYSQ 313  
QY 181 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCVNGVQWTFVHAGSXTLAGPK 240  
DB 314 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCVNGVQWTFVHAGSXTLAGPK 373  
QY 241 GPTQWYTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRDSRGSLLS 300  
DB 374 GPTQWYTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRDSRGSLLS 433  
QY 301 PRPVSYLKSSGGPILCPSSGHAAGIFRAAVCTRGVAADVPFVESHMETMR 352  
DB 434 PRPVSYLKSSGGPILCPSSGHAAGIFRAAVCTRGVAADVPFVESHMETMR 485

## RESULT 9

US-09-263-933-2  
Sequence 2, Application US/09263933

Patent No. 6280940  
GENERAL INFORMATION:

APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263, 933

CURRENT FILING DATE: 1999-03-08

EARLIER APPLICATION NUMBER: 09/129, 611

EARLIER FILING DATE: 1998-08-05

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 2307

TYPE: PRT

ORGANISM: Artificial Sequence

US-09-263-933-2

Query Match 95.4%; Score 1757; DB 3; Length 2307;

Best Local Similarity 93.8%; Pred. No. 1.1e-169;

Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILTCVHPELFDITKLLAIFGLMVLQAGITKVPYFVR 60  
DB 226 AHLHWIPLNARGRDAIILMCAVHPELFDITKLLAIFGLMVLQAGITKVPYFVR 285  
QY 61 AAGLIRACMLVRKAAAGHYVQMAFMKLAALGTYYDHLTFLQDMANAGRDIAVAEPV 120  
DB 286 AAGLIHACMLVRKAAAGHYVQMAFMKLAALGTYYDHLTFLQDMANAGRDIAVAEPV 345  
QY 121 IFSDEVKIITWGDADTAACGDIISGLPVSARKEIILGPADNFEQGMRLLAPITAYSQ 180  
DB 346 VFSDEVKIITWGDADTAACGDIISGLPVSARKEIILGPADNFEQGMRLLAPITAYSQ 405  
QY 181 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCVNGVQWTFVHAGSXTLAGPK 240  
DB 406 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCVNGVQWTFVHAGSXTLAGPK 465  
QY 241 GPTQWYTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRDSRGSLLS 300  
DB 466 GPTQWYTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRDSRGSLLS 525  
QY 301 PRPVSYLKSSGGPILCPSSGHAAGIFRAAVCTRGVAADVPFVESHMETMR 352  
DB 526 PRPVSYLKSSGGPILCPSSGHAAGIFRAAVCTRGVAADVPFVESHMETMR 577

## RESULT 10

US-09-919-901-2  
Sequence 2, Application US/09919901

Patent No. 6599738  
GENERAL INFORMATION:

APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919, 901

CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263, 933

PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129, 611

PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2

LENGTH: 2307

TYPE: PRT

ORGANISM: Artificial Sequence

US-09-919-901-2

Query Match 95.4%; Score 1757; DB 4; Length 2307;

Best Local Similarity 93.8%; Pred. No. 1.1e-169;

Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILTCVHPELFDITKLLAIFGLMVLQAGITKVPYFVR 60  
DB 226 AHLHWIPLNARGRDAIILMCAVHPELFDITKLLAIFGLMVLQAGITKVPYFVR 285

```
QY 61 AAGLIRACMLVRKAGAGHYVQMAFMKLAALTGTYYVDHLTPLODMAGLRDLAAVEPV 120
DB 286 AAGLIRACMLVRKAGAGHYVQMAFMKLGALTGYIYVHLPRLRWALAGLRDLAAVEPV 345
QY 121 IFSMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFEQGWRLLAPITAYSQ 180
DB 346 VFSMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFEQGWRLLAPITAYSQ 405
QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCWTVFHAGSKTLAGPK 240
DB 406 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCWTVFHAGSKTLAGPK 465
QY 241 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLS 300
DB 466 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLS 525
QY 301 PRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 352
DB 526 PRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 577

RESULT 11
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-11

Query Match 95.2%; Score 1754; DB 3; Length 1692;
Best Local Similarity 93.5%; Pred. No. 1,4e-169;
Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHHLOWIPIPLNVRGGDAIILITCAVHPELIDITKLLAIIPGLMYLQGITKYPYFVR 60
DB 134 AHHLOWIPIPLNVRGGDAIILITCAVHPELIDITKLLAIIPGLMYLQGITKYPYFVR 193
QY 61 AAGLIRACMLVRKAGAGHYVQMAFMKLAALTGTYYVDHLTPLODMAGLRDLAAVEPV 120
DB 194 AAGLIRACMLVRKAGAGHYVQMAFMKLGALTGYIYVHLPRLRWALAGLRDLAAVEPV 253
QY 121 IFSMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFEQGWRLLAPITAYSQ 180
DB 254 VFSMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFEQGWRLLAPITAYSQ 343
QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCWTVFHAGSKTLAGPK 240
DB 314 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCWTVFHAGSKTLAGPK 373
QY 241 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLS 300
DB 374 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLS 433
QY 301 PRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 352
DB 434 PRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 485
```

```
RESULT 12
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; EARLIER FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-11

Query Match 95.2%; Score 1754; DB 4; Length 1692;
Best Local Similarity 93.5%; Pred. No. 1,4e-169;
Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHHLOWIPIPLNVRGGDAIILITCAVHPELIDITKLLAIIPGLMYLQGITKYPYFVR 60
DB 134 AHHLOWIPIPLNVRGGDAIILITCAVHPELIDITKLLAIIPGLMYLQGITKYPYFVR 193
QY 61 AAGLIRACMLVRKAGAGHYVQMAFMKLAALTGTYYVDHLTPLODMAGLRDLAAVEPV 120
DB 194 AAGLIRACMLVRKAGAGHYVQMAFMKLGALTGYIYVHLPRLRWALAGLRDLAAVEPV 253
QY 121 IFSMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFEQGWRLLAPITAYSQ 180
DB 254 VFSMEVKIITWGDADTAACGDIISGLPVSARGREIILGPADNFEQGWRLLAPITAYSQ 313
QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCWTVFHAGSKTLAGPK 240
DB 314 QTRGLGCIITSLTGRDNQVGEVQVSTATQSFATCVNGVCWTVFHAGSKTLAGPK 373
QY 241 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLS 300
DB 374 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLS 433
QY 301 PRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 352
DB 434 PRPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 485

RESULT 13
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
```

SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2307  
; TYPE: PRF  
; ORGANISM: Artificial Sequence  
US-09-263-933-9

Query Match 95.2%; Score 1754; DB 3; Length 2307;  
Best Local Similarity 93.5%; Pred. No. 2.1e-169;  
Matches 329; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 AMLQWIPPLNVGRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 60  
DB 226 AMLQWIPPLNVGRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 285  
QY 61 AAGLIRACMLVRYKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 120  
DB 286 AAGLIRACMLVRYKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 345  
QY 121 IFSDMEVKIITWGDITACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180  
DB 346 VFSMETKRIITWGDITACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 405  
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 240  
DB 406 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 465  
QY 241 GPITQMTNVNDQDLVGWQAPPGASMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300  
DB 466 GPITQMTNVNDQDLVGWQAPPGASMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGSLLS 525  
QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFVPSMETTMR 352  
DB 526 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFVPSMETTMR 577

## RESULT 14

US-09-919-901-9  
; Sequence 9, Application US/09919901  
; Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potes, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919, 901  
; PRIOR APPLICATION NUMBER: 2001-08-02  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2307  
; TYPE: PRF  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION:  
US-09-919-901-9

Query Match 95.2%; Score 1754; DB 4; Length 2307;  
Best Local Similarity 93.5%; Pred. No. 2.1e-169;  
Matches 329; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 AMLQWIPPLNVGRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 60  
DB 226 AMLQWIPPLNVGRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 285  
QY 61 AAGLIRACMLVRYKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 120

DB 286 AAGLIRACMLVRYKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 345  
QY 121 IFSDMEVKIITWGDITACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180  
DB 346 VFSMETKRIITWGDITACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 405  
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 240  
DB 406 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 465  
QY 241 GPITQMTNVNDQDLVGWQAPPGASMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300  
DB 466 GPITQMTNVNDQDLVGWQAPPGASMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGSLLS 525  
QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFVPSMETTMR 352  
DB 526 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFVPSMETTMR 577

## RESULT 15

US-09-014-416-3  
; Sequence 3, Application US/09014416  
; Patent No. 6153421  
; GENERAL INFORMATION:  
; APPLICANT: Yanagi, Masaaki  
; APPLICANT: Burk, Jens  
; APPLICANT: Emerson, Susanne U.  
; APPLICANT: Purcell, Robert H.  
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND  
; FILE REFERENCE: 20264276  
; CURRENT APPLICATION NUMBER: US/09/014, 416  
; PRIOR FILING DATE: 1998-01-27  
; PRIOR APPLICATION NUMBER: US 60/053,062  
; EARLIER FILING DATE: 1997-07-18  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3010  
; TYPE: PRF  
; ORGANISM: Hepatitis C virus  
US-09-014-416-3

Query Match 94.9%; Score 1748; DB 3; Length 3010;  
Best Local Similarity 92.9%; Pred. No. 1.3e-168;  
Matches 327; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 AMLQWIPPLNVGRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 60  
DB 855 AMLQWIPPLNVGRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 914  
QY 61 AAGLIRACMLVRYKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 120  
DB 915 AAGLIRACMLVRYKAGGHVYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 974  
QY 121 IFSDMEVKIITWGDITACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180  
DB 975 VFSMETKRIITWGDITACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 1034  
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 240  
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 1094  
QY 241 GPITQMTNVNDQDLVGWQAPPGASMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300  
DB 1095 GPITQMTNVNDQDLVGWQAPPGASMTPTCTGSSDLYLVTRHADVIPIVRRGDSRGSLLS 1154  
QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFVPSMETTMR 352  
DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAVDFVPSMETTMR 1206

Search completed: May 6, 2004, 09:39:04

Fri May 7 13:37:03 2004

Job time : 14.8154 secs

us-10-650-585-13.ra1

Page 7



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 36.0084 Seconds  
(Without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-13

Perfect score: 1842  
Sequence: 1 AHQVWIPPLNVRGGSDAII.....RGVAKAVDPIPVESMTTWR 352

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 1140673 seqs, 27566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppa/PCF\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/PCFUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1842	100.0	352	13	US-10-017-736-13
2	1842	100.0	352	16	US-10-650-585-13
3	1842	100.0	380	13	US-10-017-736-12
4	1842	100.0	380	16	US-10-650-585-12
5	1842	100.0	393	13	US-10-017-736-11
6	1842	100.0	393	16	US-10-650-585-11
7	1842	100.0	409	13	US-10-017-736-2
8	1842	100.0	409	16	US-10-650-585-2
9	1778	96.5	341	13	US-10-017-736-14
10	1778	96.5	341	16	US-10-650-585-14
11	1766	95.9	2201	13	US-10-029-907-3
12	1766	95.9	2201	14	US-10-309-561-3
13	1766	95.9	3010	12	US-10-467-000-1
14	1757	95.4	1692	10	US-09-919-901-4
15	1757	95.4	1692	14	US-10-191-966-4

16	1757	95.4	2307	10	US-09-919-901-2	Sequence 2, Appli
17	1757	95.4	2307	14	US-10-191-966-2	Sequence 2, Appli
18	1754	95.2	1692	14	US-09-919-901-11	Sequence 11, Appl
19	1754	95.2	1692	14	US-10-191-966-11	Sequence 11, Appl
20	1754	95.2	2307	10	US-09-919-901-9	Sequence 9, Appli
21	1754	95.2	2307	14	US-10-191-966-9	Sequence 9, Appli
22	1745	94.7	1692	10	US-09-919-901-18	Sequence 18, Appl
23	1745	94.7	1692	14	US-10-191-966-18	Sequence 18, Appl
24	1745	94.7	2307	10	US-09-919-901-16	Sequence 16, Appl
25	1745	94.7	2307	14	US-10-191-966-16	Sequence 16, Appl
26	1699	92.2	2201	13	US-10-085-476-2	Sequence 2, Appli
27	1619	87.9	3011	10	US-09-742-659-4	Sequence 4, Appli
28	1619	87.9	3011	14	US-09-891-894-3	Sequence 3, Appli
29	1619	87.9	3011	14	US-10-184-150-3	Sequence 3, Appli
30	1619	87.9	3011	15	US-10-328-987-3	Sequence 3, Appli
31	1619	87.9	3012	9	US-09-238-076-2	Sequence 2, Appli
32	1619	87.9	3012	10	US-09-995-937-2	Sequence 2, Appli
33	1619	87.9	3012	10	US-09-917-563-2	Sequence 2, Appli
34	1616	87.7	3011	12	US-09-916-359-2	Sequence 2, Appli
35	1616	87.7	3011	12	US-10-296-734-406	Sequence 406, App
36	1615	87.7	3011	9	US-09-238-076-20	Sequence 20, Appli
37	1615	87.7	3011	10	US-09-995-937-20	Sequence 20, Appli
38	1615	87.7	3011	10	US-09-917-563-20	Sequence 20, Appli
39	1612	87.5	3011	9	US-09-952-572-9	Sequence 9, Appli
40	1612	87.5	3011	9	US-09-747-419-20	Sequence 20, Appli
41	1612	87.5	3011	12	US-10-189-359-14	Sequence 14, Appli
42	1612	87.5	3011	14	US-10-259-275-20	Sequence 20, Appli
43	1611	87.5	2894	9	US-09-941-611-23	Sequence 23, Appli
44	1611	87.5	2894	14	US-10-044-995-23	Sequence 23, Appli
45	1606	87.2	3011	14	US-10-232-643-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1									
US-10-017-736-13									
; Sequence 13, Application US/10017736									
; Publication No. US20020192640A1									
; GENERAL INFORMATION:									
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.									
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease									
; FILE REFERENCE: 13/082									
; CURRENT APPLICATION NUMBER: US/10/017,736									
; CURRENT FILING DATE: 2001-12-14									
; PRIOR APPLICATION NUMBER: 60/256,031									
; PRIOR FILING DATE: 2000-12-15									
; NUMBER OF SEQ ID NOS: 21									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 13									
; LENGTH: 352									
; TYPE: PRT									
; ORGANISM: HCV									
US-10-017-736-13									
Query Match									
Best Local Similarity 100.0%; Score 1842; DB 13; Length 352;									
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	1842	100.0	352	13	US-10-017-736-13	Sequence 13, Appli		
QY	2	1842	100.0	352	16	US-10-650-585-13	Sequence 16, Appli		
QY	3	1842	100.0	380	13	US-10-017-736-12	Sequence 12, Appli		
QY	4	1842	100.0	380	16	US-10-650-585-12	Sequence 12, Appli		
QY	5	1842	100.0	393	13	US-10-017-736-11	Sequence 11, Appli		
QY	6	1842	100.0	393	16	US-10-650-585-11	Sequence 11, Appli		
QY	7	1842	100.0	409	13	US-10-017-736-2	Sequence 2, Appli		
QY	8	1842	100.0	409	16	US-10-650-585-2	Sequence 2, Appli		
QY	9	1778	96.5	341	13	US-10-017-736-14	Sequence 14, Appli		
QY	10	1778	96.5	341	16	US-10-650-585-14	Sequence 14, Appli		
QY	11	1766	95.9	2201	13	US-10-029-907-3	Sequence 3, Appli		
QY	12	1766	95.9	2201	14	US-10-309-561-3	Sequence 3, Appli		
QY	13	1766	95.9	3010	12	US-10-467-000-1	Sequence 1, Appli		
QY	14	1757	95.4	1692	10	US-09-919-901-4	Sequence 4, Appli		
QY	15	1757	95.4	1692	14	US-10-191-966-4	Sequence 4, Appli		

Db 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240  
 QY 241 GPTTQMTYNDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300  
 Db 241 GPTTQMTYNDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300  
 QY 301 PRPVSYLKSGSSGGLPCPSGHAAGVIFRAAVCTRGVAKADVPIVESMETTMR 352  
 Db 301 PRPVSYLKSGSSGGLPCPSGHAAGVIFRAAVCTRGVAKADVPIVESMETTMR 352

RESULT 2

US-10-650-585-13  
 ; Sequence 13, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-13

Query Match 100.0%; Score 1842; DB 16; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 5,7e-180;  
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHLQWVPLPLNVGRGRDAIILLTCAVHPELIFDITKLLAIFGFLMWLOAGITVPEFVR 60  
 Db 1 AHLQWVPLPLNVGRGRDAIILLTCAVHPELIFDITKLLAIFGFLMWLOAGITVPEFVR 60  
 QY 61 AOGIRACMLVRKXAGGYVQMAFMKLAALGTGVYDHLTFLPDMAHAGRLDAVAEPV 120  
 Db 61 AOGIRACMLVRKXAGGYVQMAFMKLAALGTGVYDHLTFLPDMAHAGRLDAVAEPV 120  
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAETAYSQ 180  
 Db 121 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAETAYSQ 180  
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240  
 Db 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240  
 QY 241 GPTTQMTYNDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300  
 Db 241 GPTTQMTYNDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300  
 QY 301 PRPVSYLKSGSSGGLPCPSGHAAGVIFRAAVCTRGVAKADVPIVESMETTMR 352  
 Db 301 PRPVSYLKSGSSGGLPCPSGHAAGVIFRAAVCTRGVAKADVPIVESMETTMR 352

RESULT 3

US-10-017-736-12  
 ; Sequence 12, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-017-736-12

Query Match 100.0%; Score 1842; DB 13; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 6,4e-180;  
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHLQWVPLPLNVGRGRDAIILLTCAVHPELIFDITKLLAIFGFLMWLOAGITVPEFVR 60  
 Db 29 AHLQWVPLPLNVGRGRDAIILLTCAVHPELIFDITKLLAIFGFLMWLOAGITVPEFVR 88  
 QY 61 AOGIRACMLVRKXAGGYVQMAFMKLAALGTGVYDHLTFLPDMAHAGRLDAVAEPV 120  
 Db 89 AOGIRACMLVRKXAGGYVQMAFMKLAALGTGVYDHLTFLPDMAHAGRLDAVAEPV 148  
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAETAYSQ 180  
 Db 149 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAETAYSQ 208  
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240  
 Db 209 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 268  
 QY 241 GPTTQMTYNDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300  
 Db 269 GPTTQMTYNDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 328  
 QY 301 PRPVSYLKSGSSGGLPCPSGHAAGVIFRAAVCTRGVAKADVPIVESMETTMR 352  
 Db 329 PRPVSYLKSGSSGGLPCPSGHAAGVIFRAAVCTRGVAKADVPIVESMETTMR 380

RESULT 4

US-10-650-585-12  
 ; Sequence 12, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-12

Query Match 100.0%; Score 1842; DB 16; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 6,4e-180;  
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHLQWVPLPLNVGRGRDAIILLTCAVHPELIFDITKLLAIFGFLMWLOAGITVPEFVR 60  
 Db 29 AHLQWVPLPLNVGRGRDAIILLTCAVHPELIFDITKLLAIFGFLMWLOAGITVPEFVR 88  
 QY 61 AOGIRACMLVRKXAGGYVQMAFMKLAALGTGVYDHLTFLPDMAHAGRLDAVAEPV 120  
 Db 89 AOGIRACMLVRKXAGGYVQMAFMKLAALGTGVYDHLTFLPDMAHAGRLDAVAEPV 148  
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAETAYSQ 180



```

Db 149 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGADNFEQGMWLLAPITAYSQ 208
Qy 181 QTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCNVGVMTVFHAGSKTLAGPK 240
Db 209 QTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCNVGVMTVFHAGSKTLAGPK 268
Qy 241 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
Db 269 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 328
Qy 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352
Db 329 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 380

```

## RESULT 5

```

US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

```

```

Query Match 100.0%; Score 1842; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,7e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AHQWIPPLNVRGGSDAIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 60
Db 42 AHQWIPPLNVRGGSDAIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 101
Qy 61 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPQDMAHAGRLDAVAVEPV 120
Db 102 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPQDMAHAGRLDAVAVEPV 161
Qy 121 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGADNFEQGMWLLAPITAYSQ 180
Db 162 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGADNFEQGMWLLAPITAYSQ 221
Qy 181 QTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCNVGVMTVFHAGSKTLAGPK 240
Db 222 QTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCNVGVMTVFHAGSKTLAGPK 281
Qy 241 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
Db 282 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 341
Qy 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352
Db 342 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 393

```

## RESULT 6

```

US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082

```

```

; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

```

```

Query Match 100.0%; Score 1842; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,7e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AHQWIPPLNVRGGSDAIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 60
Db 42 AHQWIPPLNVRGGSDAIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 101
Qy 61 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPQDMAHAGRLDAVAVEPV 120
Db 102 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPQDMAHAGRLDAVAVEPV 161
Qy 121 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGADNFEQGMWLLAPITAYSQ 180
Db 162 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGADNFEQGMWLLAPITAYSQ 221
Qy 181 QTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCNVGVMTVFHAGSKTLAGPK 240
Db 222 QTRGLGCIITSLTGRDKQVEGEVQVSTATOSFLATCNVGVMTVFHAGSKTLAGPK 281
Qy 241 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
Db 282 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 341
Qy 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352
Db 342 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 393

```

## RESULT 7

```

US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2

```

```

Query Match 100.0%; Score 1842; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 7,1e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AHQWIPPLNVRGGSDAIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 60
Db 46 AHQWIPPLNVRGGSDAIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 105
Qy 61 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPQDMAHAGRLDAVAVEPV 120

```

Db 106 AAGIIRACMLVRKAGGHVYQMAFMKLAALTGVYVDHLTPLODMAHAGRLDAVAVEPV 165  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-2

Qy 121 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNPEGQWELAPITAYQ 180  
Db 166 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNPEGQWELAPITAYQ 225  
Qy 181 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFVFGAGSKTLAGPK 240  
Db 226 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFVFGAGSKTLAGPK 285  
Qy 241 GPIQWMTNVDDQVLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVRRGDSRGLS 300  
Db 286 GPIQWMTNVDDQVLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVRRGDSRGLS 345  
Qy 301 PRPVSYLKSSGGPILCPSGHANGIFPRAAVCTRGVAKAVDFIVESMETMR 352  
Db 346 PRPVSYLKSSGGPILCPSGHANGIFPRAAVCTRGVAKAVDFIVESMETMR 397

## RESULT 8

US-10-650-585-2  
; Sequence 2, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-2

Query Match 100.0%; Score 1842; DB 16; Length 409;  
Best Local Similarity 100.0%; Pred. No. 7,1e-180;  
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AALQWIPPLNVGRGRDAIILLTCVAHPPELLFDITKLLAIFGPIMLVLAQGITKVEYFVR 60  
Db 46 AALQWIPPLNVGRGRDAIILLTCVAHPPELLFDITKLLAIFGPIMLVLAQGITKVEYFVR 105  
Qy 61 AAGIIRACMLVRKAGGHVYQMAFMKLAALTGVYVDHLTPLODMAHAGRLDAVAVEPV 120  
Db 106 AAGIIRACMLVRKAGGHVYQMAFMKLAALTGVYVDHLTPLODMAHAGRLDAVAVEPV 165  
Qy 121 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNPEGQWELAPITAYQ 180  
Db 166 IFSDMEVKIITWGADTAACGDIISGLPVSARRGREILGPADNPEGQWELAPITAYQ 225  
Qy 181 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFVFGAGSKTLAGPK 240  
Db 226 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFVFGAGSKTLAGPK 285  
Qy 241 GPIQWMTNVDDQVLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVRRGDSRGLS 300  
Db 286 GPIQWMTNVDDQVLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVRRGDSRGLS 345  
Qy 301 PRPVSYLKSSGGPILCPSGHANGIFPRAAVCTRGVAKAVDFIVESMETMR 352  
Db 346 PRPVSYLKSSGGPILCPSGHANGIFPRAAVCTRGVAKAVDFIVESMETMR 397

RESULT 9  
US-10-017-736-14  
; Sequence 14, Application US/10017736

; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-017-736-14

Query Match 96.5%; Score 1778; DB 13; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2e-173;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 VRGGRDAIILLTCVAHPPELLFDITKLLAIFGPIMLVLAQGITKVEYFVRAGILRACMLV 71  
Db 1 VRGGRDAIILLTCVAHPPELLFDITKLLAIFGPIMLVLAQGITKVEYFVRAGILRACMLV 60  
Qy 72 RKAAGGHVYQMAFMKLAALTGVYVDHLTPLODMAHAGRLDAVAVEPVIFSDMEVKIIT 131  
Db 61 RKAAGGHVYQMAFMKLAALTGVYVDHLTPLODMAHAGRLDAVAVEPVIFSDMEVKIIT 120  
Qy 122 WCADTAACGDIISGLPVSARRGREILGPADNPEGQWELAPITAYSOQTRGLACIIT 191  
Db 121 WCADTAACGDIISGLPVSARRGREILGPADNPEGQWELAPITAYSOQTRGLACIIT 180  
Qy 192 SLTGRKNOVEGEVQVSTATOSFLATCNVCWTFVFGAGSKTLAGPKGITQMTYND 251  
Db 181 SLTGRKNOVEGEVQVSTATOSFLATCNVCWTFVFGAGSKTLAGPKGITQMTYND 240  
Qy 252 QDLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVRRGDSRGLSPRPVSYLKSS 311  
Db 241 QDLVGMQAPPGARSMTCTCGSSDLVYTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300  
Qy 312 GGPILCPSGHANGIFPRAAVCTRGVAKAVDFIVESMETMR 352  
Db 301 GGPILCPSGHANGIFPRAAVCTRGVAKAVDFIVESMETMR 341

RESULT 10  
US-10-650-585-14  
; Sequence 14, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; PRIOR FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017,736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-14

Query Match 96.5%; Score 1778; DB 16; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2e-173;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 VRGGRDAIILLTCVAHPPELLFDITKLLAIFGPIMLVLAQGITKVEYFVRAGILRACMLV 71

Db 1 VRGRDAIILITCAVHPELIFDITKLLALFPGPLMTVQAGITKVPYVRAQGLIRACMLV 60

Qy 72 RKAAGHYVQMAFMKLAALITGYTYYDLTLPLOMAHAGLDLAVAVEPVFSPMEVKIT 131

Db 61 RKAAGHYVQMAFMKLAALITGYTYYDLTLPLOMAHAGLDLAVAVEPVFSPMEVKIT 120

Qy 132 WGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSOQTRGLGCIIT 191

Db 121 WGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSOQTRGLGCIIT 180

Qy 192 SLTRDKNQVGEVQVAVSTATOSFLATCVNGVCTVFGAGSKTLAPKSPITQMTYND 251

Db 181 SLTRDKNQVGEVQVAVSTATOSFLATCVNGVCTVFGAGSKTLAPKSPITQMTYND 240

Qy 252 QDLVGMQAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGRSGSLSPRVSYLKSS 311

Db 241 QDLVGMQAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGRSGSLSPRVSYLKSS 300

Qy 312 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETTR 352

Db 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETTR 341

## RESULT 11

US-10-029-907-3

Sequence 3, Application US/10029907

Publication No. US20020142350A1

GENERAL INFORMATION:

APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.

TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM

FILE REFERENCE: 13/083

CURRENT APPLICATION NUMBER: US/10/029,907

PRIOR FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 60/257,857

PRIOR FILING DATE: 2000-12-22

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 2201

TYPE: PR

ORGANISM: HCV

FEATURE:

NAME/KEY: VARIANT

LOCATION: 882

OTHER INFORMATION: Xaa is Lys or Arg

NAME/KEY: VARIANT

LOCATION: 1489

OTHER INFORMATION: Xaa is Leu

US-10-029-907-3

Query Match 95.9%; Score 1766; DB 13; Length 2201;

Best Local Similarity 94.3%; Pred. No. 4,7e-171;

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

Qy 1 AHLQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLALFPGPLMTVQAGITKVPYFVR 60

Db 46 AHLQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLALFPGPLMTVQAGITKVPYFVR 105

Qy 61 AAGLIRACMLVRKAGHYVQMAFMKLAALITGYTYYDLTLPLOMAHAGLDLAVAVEPV 120

Db 106 AAGLIRACMLVRKAGHYVQMAFMKLAALITGYTYYDLTLPLOMAHAGLDLAVAVEPV 165

Qy 121 IFSMEVKITWGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSO 180

Db 166 VFSMEVKITWGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSO 225

Qy 181 QTRGLGCIITSLTRDKNQVGEVQVAVSTATOSFLATCVNGVCTVFGAGSKTLAPK 240

Db 226 QTRGLGCIITSLTRDKNQVGEVQVAVSTATOSFLATCVNGVCTVFGAGSKTLAPK 285

Qy 241 GPITQMTYNDQDLVGMQAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGRSGSLIS 300

Db 286 GPITQMTYNDQDLVGMQAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGRSGSLIS 345

Qy 301 PRPVSYLKSSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETTR 352

Db 346 PRPVSYLKSSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETTR 397

## RESULT 12

US-10-309-561-3

Sequence 3, Application US/10309561

Publication No. US20030148348A1

GENERAL INFORMATION:

APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.

TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM

FILE REFERENCE: 13/083

CURRENT APPLICATION NUMBER: US/10/309,561

PRIOR FILING DATE: 2002-12-04

PRIOR APPLICATION NUMBER: US/10/029,907

PRIOR FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 60/257,857

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 2201

TYPE: PR

ORGANISM: HCV

FEATURE:

NAME/KEY: VARIANT

LOCATION: 882

OTHER INFORMATION: Xaa is Lys or Arg

NAME/KEY: VARIANT

LOCATION: 1489

OTHER INFORMATION: Xaa is Leu

US-10-309-561-3

Query Match 95.9%; Score 1766; DB 14; Length 2201;

Best Local Similarity 94.3%; Pred. No. 4,7e-171;

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

Qy 1 AHLQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLALFPGPLMTVQAGITKVPYFVR 60

Db 46 AHLQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLALFPGPLMTVQAGITKVPYFVR 105

Qy 61 AAGLIRACMLVRKAGHYVQMAFMKLAALITGYTYYDLTLPLOMAHAGLDLAVAVEPV 120

Db 106 AAGLIRACMLVRKAGHYVQMAFMKLAALITGYTYYDLTLPLOMAHAGLDLAVAVEPV 165

Qy 121 IFSMEVKITWGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSO 180

Db 166 VFSMEVKITWGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSO 225

Qy 181 QTRGLGCIITSLTRDKNQVGEVQVAVSTATOSFLATCVNGVCTVFGAGSKTLAPK 240

Db 226 QTRGLGCIITSLTRDKNQVGEVQVAVSTATOSFLATCVNGVCTVFGAGSKTLAPK 285

Qy 301 PRPVSYLKSSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETTR 352

Db 346 PRPVSYLKSSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETTR 397

## RESULT 13

US-10-467-000-1

Sequence 1, Application US/10467000

Publication No. US20040067486A1

GENERAL INFORMATION:

```

; APPLICANT: De Francesco, Raffaele
; APPLICANT: Magliacchio, Giovanni
; APPLICANT: Pavesa, Giacomo
; TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
; FILE OF INVENTION: ENHANCED CELLS
; FILE REFERENCE: ITR0003P
; CURRENT APPLICATION NUMBER: US/10/467,000
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: PCT/EP02/00526
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/263,479
; PRIOR FILING DATE: 2001-01-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Con 1 HCV isolate nucleic acid
; US-10-467-000-1

```

```

Query Match          95.4%; Score 1766; DB 12; Length 3010;
Best Local Similarity 94.3%; Pred. No. 7,4e-171;
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 AHHQWIPPLNVRGGRDAIILTCVAPHELIFDITKLLAIFGPMVLQAGITKVPYFVR 60
DB AHHQWIPPLNVRGGRDAIILTCVAPHELIFDITKLLAIFGPMVLQAGITKVPYFVR 914
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 120
DB AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 974
QY 915 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 974
DB AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 974
QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180
DB 975 VFSMEKVIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 1034
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCMVTFHAGSXTLAGPK 240
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCMVTFHAGSXTLAGPK 1094
QY 241 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGRDSRGLS 300
DB 1095 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGRDSRGLS 1154
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESHMETTR 352
DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESHMETTR 1206

```

```

RESULT 14
US-09-919-901-4
; Sequence 4, Application US/09919901
; Publication No. US20030082518A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION:
; US-09-919-901-4

```

```

Query Match          95.4%; Score 1757; DB 10; Length 1692;
Best Local Similarity 93.8%; Pred. No. 2.7e-170;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 AHHQWIPPLNVRGGRDAIILTCVAPHELIFDITKLLAIFGPMVLQAGITKVPYFVR 60
DB AHHQWIPPLNVRGGRDAIILTCVAPHELIFDITKLLAIFGPMVLQAGITKVPYFVR 193
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 120
DB 194 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 253
QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180
DB 254 VFSMEKVIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 313
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCMVTFHAGSXTLAGPK 240
DB 314 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVCMVTFHAGSXTLAGPK 373
QY 241 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGRDSRGLS 300
DB 374 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGRDSRGLS 433
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESHMETTR 352
DB 434 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESHMETTR 485

```

```

RESULT 15
US-10-191-966-4
; Sequence 4, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; US-10-191-966-4

```

```

Query Match          95.4%; Score 1757; DB 14; Length 1692;
Best Local Similarity 93.8%; Pred. No. 2.7e-170;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 AHHQWIPPLNVRGGRDAIILTCVAPHELIFDITKLLAIFGPMVLQAGITKVPYFVR 60
DB 134 AHHQWIPPLNVRGGRDAIILTCVAPHELIFDITKLLAIFGPMVLQAGITKVPYFVR 193
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 120
DB 194 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 253
QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180
DB 194 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLLPLQDMAHAGRLDAVAPEV 253

```

```

Db      254 VESDMETKITWADTAACDIIILGLPVASARGKEIILGPADSIEGRGMRLLAPTAYSQ 313
QY      181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFLATCVNGVCMVTFHAGSKTLAGPK 240
Db      314 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFLATCVNGVCMVTFHAGSKTLAGPK 373
QY      241 GPITQMTNVDQDIVGQAPPGARSMTPTCTGSSDLVYVTRHADVIPIVRRRGSRLS 300
Db      374 GPITQMTNVDQDIVGQAPPGARSMTPTCTGSSDLVYVTRHADVIPIVRRRGSRLS 433
QY      301 PRPVSYLKSSSGPPLCPSGHVGIFRAAVCTRGVAKAVDFIPVESMETMR 352
Db      434 PRPVSYLKSSSGPPLCPSGHVGIFRAAVCTRGVAKAVDFIPVESMETMR 485

```

Search completed: May 6, 2004, 09:43:21  
 Job time : 37.0084 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 / Search time 46.4159 Seconds

(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-14  
Perfect score: 1778  
Sequence: 1 VRRGRDAIILLTCVAPHELI.....RGVAKAVDFPVSMTETTR 341

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	1778	100.0	341	5	ABG32187 HCV prote
2	1778	100.0	352	5	ABG32186 HCV prote
3	1778	100.0	380	5	ABG32185 HCV prote
4	1778	100.0	393	5	ABG32184 HCV prote
5	1778	100.0	409	5	ABG32181 HCV prote
6	1710	96.2	3010	2	AAR82694 Partial H
7	1705	95.9	3010	2	AAR86822 HCV prote
8	1704	95.8	3010	2	AAR68864 Hepatitis
9	1702	95.7	2201	5	ABG30591 Hepatitis
10	1702	95.7	2201	5	ABG30591 Hepatitis
11	1702	95.7	2201	5	ABG30600 Hepatitis
12	1702	95.7	2201	5	ABG30581 Hepatitis
13	1702	95.7	2201	5	ABG30593 Hepatitis
14	1702	95.7	2201	5	ABG30582 Hepatitis
15	1702	95.7	2201	5	ABG30580 Hepatitis
16	1702	95.7	2201	5	ABG30587 Hepatitis
17	1702	95.7	2201	5	ABG30599 Hepatitis
18	1702	95.7	2201	5	ABG30594 Hepatitis
19	1702	95.7	2201	5	ABG30598 Hepatitis
20	1702	95.7	3010	5	ABG30595 Hepatitis
21	1702	95.7	3010	5	ABG32458 Hepatitis
22	1702	95.7	3010	5	ABG32459 Hepatitis
23	1702	95.7	3010	5	ABG32451 Hepatitis
24	1702	95.7	3010	5	ABG32455 Hepatitis
25	1702	95.7	3010	5	ABG32457 Hepatitis

26	1702	95.7	3010	5	ABG32460 Hepatitis
27	1702	95.7	3010	5	ABG32461 Hepatitis
28	1702	95.7	3010	5	ABG32454 Hepatitis
29	1702	95.7	3011	5	ABG32456 Hepatitis
30	1699	95.6	2201	5	ABG30586 Hepatitis
31	1699	95.6	2201	5	ABG30589 Hepatitis
32	1699	95.6	2201	5	ABG30583 Hepatitis
33	1699	95.6	2201	5	ABG30588 Hepatitis
34	1698	95.5	2201	5	ABG30590 Hepatitis
35	1698	95.5	2307	3	AAV70064 Recombina
36	1697	95.4	3010	5	AAE20477 HCV-S1 fu
37	1696	95.4	3010	5	ABG32452 Recombina
38	1696	95.3	768	2	AAR40223 Recombina
39	1695	95.3	2201	5	ABG30584 Hepatitis
40	1695	95.3	2201	5	ABG30602 Recombina
41	1695	95.3	2307	3	AAV70065 Recombina
42	1695	95.3	3010	5	ABG32453 Hepatitis
43	1691	95.1	2201	5	ABG30585 Hepatitis
44	1688	94.9	3014	2	AAR54099 NANBV E1
45	1687	94.9	3010	2	AAW98022 Infectiou

## ALIGNMENTS

RESULT 1	ABG32187	standard; protein; 341 AA.
ID	ABG32187	
AC	ABG32187;	
XX		
AC	ABG32187;	
XX		
DT	05-NOV-2002	(first entry)
XX		
DE	HCV protease NS2/3 truncation mutant 866-1206.	
XX		
KM	HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;	
KM	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;	
KM	hepatocellular carcinoma; anti-inflammatory; lauryldiethylamine oxide; LDAO;	
KM	chaotropic agent; mutant; mutagen.	
XX		
OS	Hepatitis C virus.	
OS	Synthetic.	
XX		
PN	W0200248375-A2.	
XX		
PD	20-UTN-2002.	
XX		
PF	13-DEC-2001; 2001W0-CA001796.	
XX		
PR	15-DEC-2000; 2000US-0256031P.	
XX		
PA	(BOEHR) BOEHRINGER INGELHEIM CANADA LTD.	
XX		
PI	Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;	
XX	WPI; 2002-599511/64.	
DR		
XX		
PT	Novel polypeptide for screening inhibitors of non-structural proteases	
PT	useful as therapeutic agents against hepatitis C virus, comprises full	
PT	length non-structural protease, or its truncation.	
XX		
PS	Claim 41; Page 62-63; 67pp; English.	
XX		
CC	The invention relates to an isolated polypeptide consisting of a full-	
CC	length HCV (hepatitis C virus) non-structural (NS2/3) protease (referred	
CC	to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal	
CC	residue amino acid 810 to 906, or having a minimal amino acid sequence	
CC	from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length	
CC	NS2/3 protease. Also included are (1) a composition (C) comprising an	
CC	isolated HCV NS2/3 protease selected from full length NS2/3 protease, or	
CC	its truncation or a mutated sequence, where the protease is in a solution	
CC	comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)	
CC	to prevent auto-cleavage of the protease, (2) a NS2/3 inhibitory peptide	

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 866-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 341 AA:

Query Match 100.0%; Score 1778; DB 5; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 9,6e-164;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCVHPELFDITKLLAFGPIMLVQAGITVPYVRAQGLIRACMLV 60  
 DB 1 VRGRDAIILLTCVHPELFDITKLLAFGPIMLVQAGITVPYVRAQGLIRACMLV 60  
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPQDMNAHGLDLVAVEPVFSDEMEVKIT 120  
 DB 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPQDMNAHGLDLVAVEPVFSDEMEVKIT 120  
 QY 121 WGADTAAAGDITISGLPVARSRGREILIGPADNFEQGRLLAPITAVSQOTRGLGCIIT 180  
 DB 121 WGADTAAAGDITISGLPVARSRGREILIGPADNFEQGRLLAPITAVSQOTRGLGCIIT 180  
 QY 121 WGADTAAAGDITISGLPVARSRGREILIGPADNFEQGRLLAPITAVSQOTRGLGCIIT 180  
 DB 121 WGADTAAAGDITISGLPVARSRGREILIGPADNFEQGRLLAPITAVSQOTRGLGCIIT 180  
 QY 181 SLTRGDKNOYGEVQVSTATQSPATCVCVCTVHGAGSKTLGPKPIQOMYTNVD 240  
 DB 181 SLTRGDKNOYGEVQVSTATQSPATCVCVCTVHGAGSKTLGPKPIQOMYTNVD 240  
 QY 241 QDLVGMQAPPCARSMTPTCTGSSDLYLVTNRADYIPVRRGDSRGSILSPVSYLKSS 300  
 DB 241 QDLVGMQAPPCARSMTPTCTGSSDLYLVTNRADYIPVRRGDSRGSILSPVSYLKSS 300  
 QY 301 GGPILLCSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTNR 341  
 DB 301 GGPILLCSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTNR 341

RESULT 2  
 ABG32186  
 ID ABG32186 standard; protein; 352 AA.

XX ABG32186;  
 AC 05-NOV-2002 (first entry)  
 DT HCV protease NS2/3 truncation mutant 855-1206.  
 DB HCV protease NS2/3 truncation mutant 855-1206.  
 XX HCV, enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
 KW chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KW hepatocellular carcinoma; hepatocellular carcinoma; hepatocellular carcinoma;  
 KW hepatocellular carcinoma; hepatocellular carcinoma; hepatocellular carcinoma;  
 KW hepatocellular carcinoma; hepatocellular carcinoma; hepatocellular carcinoma;  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX

PN WO200248375-A2.  
 XX 20-JUN-2002.  
 PD 13-DEC-2001; 2001WO-C0001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 PR (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 PI WPI; 2002-599511/64.  
 DR Novel polypeptide for screening inhibitors of non-structural proteases  
 XX useful as therapeutic agents against hepatitis C virus, comprises full  
 XX length non-structural protease, or its truncation.  
 XX Claim 41; Page 61-62; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 855-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 352 AA:

Query Match 100.0%; Score 1778; DB 5; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 1e-163;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCVHPELFDITKLLAFGPIMLVQAGITVPYVRAQGLIRACMLV 60  
 DB 12 VRGRDAIILLTCVHPELFDITKLLAFGPIMLVQAGITVPYVRAQGLIRACMLV 71  
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPQDMNAHGLDLVAVEPVFSDEMEVKIT 120  
 DB 72 RKAAGHYVQMAFMKLAALGTYYVDHLTPQDMNAHGLDLVAVEPVFSDEMEVKIT 131  
 QY 121 WGADTAAAGDITISGLPVARSRGREILIGPADNFEQGRLLAPITAVSQOTRGLGCIIT 180  
 DB 132 WGADTAAAGDITISGLPVARSRGREILIGPADNFEQGRLLAPITAVSQOTRGLGCIIT 191  
 QY 181 SLTRGDKNOYGEVQVSTATQSPATCVCVCTVHGAGSKTLGPKPIQOMYTNVD 240



|||||  
DB SLTRDKNQVEGVAVSTATQSLATCAVGCWTFVHGAGSKTLAPKPIQTMTND 251  
QY 192 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVPVRRRGRSGSLSPRVSYLKGS 300  
DB 241 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVPVRRRGRSGSLSPRVSYLKGS 311  
QY 252 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVPVRRRGRSGSLSPRVSYLKGS 311  
DB 301 GGPLCPSGHAIVGIFRAVCTRGVAKAVDFIVESMETMR 341  
QY 312 GGPLCPSGHAIVGIFRAVCTRGVAKAVDFIVESMETMR 352

RESULT 3  
ABG32185  
ID ABG32185 standard; protein; 380 AA.  
XX  
AC ABG32185;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 827-1206.  
XX  
KM HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
KM chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN W0200248375-A2.  
XX  
PD 20-JUN-2002.  
XX  
PF 13-DEC-2001; 2001WO-CA001796.  
XX  
PR 15-DEC-2000; 2000US-0256031P.  
XX  
PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX  
DR WPI; 2002-599511/64.  
PT Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.  
XX  
PS Claim 41; Page 60-61; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an

CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation mutant 827-1206  
CC (numbered relative to the full length NS2/3 protein)  
XX  
SQ Sequence 380 AA;  
XX

Query Match 100.0%; Score 1778; DB 5; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1,1e-163;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VRGGRDAIILLTCANVHPELIPDITKLLAIFGPLVWLQAGITKVPYFPAQGLIRACMLV 60  
DB 40 VRGGRDAIILLTCANVHPELIPDITKLLAIFGPLVWLQAGITKVPYFPAQGLIRACMLV 99  
QY 61 RRAAGGHVYQMAFMKLAALTGTYYVDHLTPLODMWAGLRDLAVAVEPIFSDMEVKIT 120  
DB 100 RRAAGGHVYQMAFMKLAALTGTYYVDHLTPLODMWAGLRDLAVAVEPIFSDMEVKIT 159  
QY 121 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGWELLAPITAYSQOTRGLGCIIT 180  
DB 160 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGWELLAPITAYSQOTRGLGCIIT 219  
QY 181 SLTGRDKNQVEGVAVSTATQSLATCAVGCWTFVHGAGSKTLAPKPIQTMTND 240  
DB 220 SLTGRDKNQVEGVAVSTATQSLATCAVGCWTFVHGAGSKTLAPKPIQTMTND 279  
QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVPVRRRGRSGSLSPRVSYLKGS 300  
DB 280 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVPVRRRGRSGSLSPRVSYLKGS 339  
QY 301 GGPLCPSGHAIVGIFRAVCTRGVAKAVDFIVESMETMR 341  
DB 340 GGPLCPSGHAIVGIFRAVCTRGVAKAVDFIVESMETMR 380

RESULT 4  
ABG32184  
ID ABG32184 standard; protein; 393 AA.  
XX  
AC ABG32184;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 815-1206.  
XX  
KM HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
KM chronic liver disease; cirrhosis; end-stage liver disease; virucide;  
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN W0200248375-A2.  
XX  
PD 20-JUN-2002.  
XX  
PF 13-DEC-2001; 2001WO-CA001796.  
XX  
PR 15-DEC-2000; 2000US-0256031P.  
XX  
PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX  
DR WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PT length non-structural protease, or its truncation.

PS Claim 41; Page 59-60; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterization of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation mutant 815-1206  
CC (numbered relative to the full length NS2/3 protein)

XX Sequence 393 AA:

SO Query Match 100.0%; Score 1778; DB 5; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1,2e-163;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGRDAILLLTCAVHPELIFDIKLLALFGLPLVLAQGITVPPFVRAQGLIRACMY 60  
Db 53 VRGRDAILLLTCAVHPELIFDIKLLALFGLPLVLAQGITVPPFVRAQGLIRACMY 112

QY 61 RKAAGHYVQWAFKALATCTYTDLTPLQWMAHGLDLAVNEPVFSDMEVKIIT 120  
Db 113 RKAAGHYVQWAFKALATCTYTDLTPLQWMAHGLDLAVNEPVFSDMEVKIIT 172

QY 121 WGAADTAACGDIISGLPVSARGRRIILGPADNFGQGMRLAPITAVSQOTRLGCIIT 180  
Db 173 WGAADTAACGDIISGLPVSARGRRIILGPADNFGQGMRLAPITAVSQOTRLGCIIT 232

QY 181 SLTRDRNQVGEVQVSTATQSHFLATCVAGVCTVPHGAGSTLAPKPIITQMTTND 240  
Db 233 SLTRDRNQVGEVQVSTATQSHFLATCVAGVCTVPHGAGSTLAPKPIITQMTTND 292

QY 241 QDLWGMQAPPGARSMPTCTGSSDLVYVTHADVIPRRAGDSRGLISRPVSYKGS 300  
Db 293 QDLWGMQAPPGARSMPTCTGSSDLVYVTHADVIPRRAGDSRGLISRPVSYKGS 352

QY 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMETTNR 341  
Db 353 GGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMETTNR 393

\* RESULT 5

ABG32181  
ID ABG32181 standard; protein; 409 AA.  
XX  
XX AC ABG32181;  
XX  
XX DT 05-NOV-2002 (first entry)  
XX  
XX DE HCV protease NS2/3 (810-1206).  
XX  
XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
XX chronic liver disease; cirrhosis; end-stage liver disease; viremia;  
XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
XX chaotropic agent; mutant; mutein.  
XX  
XX OS Hepatitis C virus.  
XX Synthetic.  
FH Key Location/Qualifiers  
FT Peptide 398..409  
FT /note="Streptavidin tag"  
XX  
XX WO200248375-A2.  
XX  
XX PD 20-JUN-2002.  
XX  
XX PF 13-DEC-2001; 2001WO-CA001796.  
XX  
XX PR 15-DEC-2000; 2000US-0256031P.  
XX  
XX PA (BOE ) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
XX PI Thibeault D, Lemaire D, Maurice R, Pilote L, Pause A;  
XX  
XX DR WPI; 2002-599511/64.  
XX  
XX DR N-PSDB; ABK90406.  
XX  
XX PT Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX  
XX Claim 42; Fig 1B; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
CC protease, involving isolating the protease in the presence of a  
CC chaotropic agent, refolding the isolated protease by contacting it with a  
CC reducing agent, and LDAO in the presence of reduced concentration of the  
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
CC containing an activation detergent to induce auto-cleavage of the NS2/3  
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
CC protease, involving incubating the active NS2/3 protease produced by M2  
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
CC cleavage products or their fragments, and measuring the presence or  
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterization of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The

CC present sequence represents the NS2/3 (810-1206) protein, which has a C-terminal streptavidin tag

XX Sequence 409 AA;

Query Match 100.0%; Score 1778; DB 5; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1,2e-163; Indels 0; Gaps 0;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 60  
Db VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 116

QY 61 RKAAGHYVQMAFMKLAALITGYVDHITPLQDMAHAGIRDLAFAVEPVFSDMEVKIIT 120  
Db RKAAGHYVQMAFMKLAALITGYVDHITPLQDMAHAGIRDLAFAVEPVFSDMEVKIIT 176

QY 121 WGAADTAACGDIISGLPVSAARRREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180  
Db WGAADTAACGDIISGLPVSAARRREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 236

QY 181 SLTRGRDNQVGEVQVYSTATQSLFATCVNGCMTVFHAGSKTLAEPKPIITOMYTNVD 240  
Db SLTRGRDNQVGEVQVYSTATQSLFATCVNGCMTVFHAGSKTLAEPKPIITOMYTNVD 296

QY 241 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSRGSLSPPVSYLKSS 300  
Db QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSRGSLSPPVSYLKSS 356

QY 301 GGPILCPSGHNAVGIFFRAVCTRGVAKAVDFIVESMETTMR 341  
Db GGPILCPSGHNAVGIFFRAVCTRGVAKAVDFIVESMETTMR 397

RESULT 6  
AAR82694  
ID AAR82694 standard; protein; 3010 AA.

AC AAR82694;  
DT 16-OCT-2003 (revised)  
DT 14-NOV-1996 (first entry)

DE Partial HCV non-structural polyprotein.

KW proteinase; hepatitis C virus; screening; inhibitor; proteolytic;  
KW identification; cleavage.

OS Hepatitis C virus; Virus.

XX Key Location/Qualifiers

FT Protein 898..1233  
FT /note="partial proteinase; see AAR82692"  
FT 992..1907  
FT /note="partial proteinase; see AAR82693"

XX JF07184648-A.

XX 25-JUL-1995.

XX 05-FEB-1993; 93JP-00018854.

XX 07-FEB-1992; 92JP-00022657.

XX 18-SEP-1992; 92JP-00249240.

XX 04-DEC-1992; 92JP-00325303.

XX (KAEN/) KAENNO K.  
XX (SUMO) SUMITOMO METAL IND LTD.  
XX (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

XX WPI; 1995-287962/38.  
XX N-PSDB; AAT03960.

PT An HCV proteinase active substance - which has activity as an anti-HCV agent and can be used to screen for proteinase inhibitors.

XX Disclosure; Page 39-48; 52pp; Japanese.

XX The present sequence is a partial Hepatitis C Virus (HCV) polyprotein from the non-structural region. Partial proteinase sequences (AAR82692-93) are contained within this sequence. The proteinases can be used as anti-HCV agents. They can also be used to screen cdds. for their ability to inhibit their proteolytic activity. In this way proteinase inhibitors can be identified. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 3010 AA;

Query Match 96.2%; Score 1710; DB 2; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 8.3e-156; Indels 0; Gaps 0;  
Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 60  
Db VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 925

QY 61 RKAAGHYVQMAFMKLAALITGYVDHITPLQDMAHAGIRDLAFAVEPVFSDMEVKIIT 120  
Db RKAAGHYVQMAFMKLAALITGYVDHITPLQDMAHAGIRDLAFAVEPVFSDMEVKIIT 985

QY 121 WGAADTAACGDIISGLPVSAARRREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180  
Db WGAADTAACGDIISGLPVSAARRREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 1045

QY 181 SLTRGRDNQVGEVQVYSTATQSLFATCVNGCMTVFHAGSKTLAEPKPIITOMYTNVD 240  
Db SLTRGRDNQVGEVQVYSTATQSLFATCVNGCMTVFHAGSKTLAEPKPIITOMYTNVD 1105

QY 241 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSRGSLSPPVSYLKSS 300  
Db QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSRGSLSPPVSYLKSS 1165

QY 301 GGPILCPSGHNAVGIFFRAVCTRGVAKAVDFIVESMETTMR 341  
Db GGPILCPSGHNAVGIFFRAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 7  
AAR68622  
ID AAR68622 standard; protein; 3010 AA.  
XX AAR68622;

DT 16-OCT-2003 (revised)  
DT 16-OCT-1995 (first entry)

DE HCV protein cleavable with new serine proteinase.

KW proteinase; serine; cleavage; hepatitis C virus; HCV.

OS Hepatitis C virus; Virus.

XX Key Location/Qualifiers

FT Cleavage-site 2419..2420  
FT /note="serine protease cleavage site"

XX JF06315377-A.

XX 15-NOV-1994.

XX 06-MAY-1993; 93JP-00105666.

XX 06-MAY-1993; 93JP-00105666.

XX (KAEN/) KAENNO K.  
XX (SUMO) SUMITOMO METAL IND LTD.  
XX (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

XX	WP1: 1995-032330/05.
DR	N-FSDB: AAQ80498.
XX	
PT	New HCV-originated proteinase active substance - used for site-specific cleavage by an intermolecular reaction and the purification thereof.
PS	Disclosure; Page 10-19; 23pp; Japanese.
CC	This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is cleaved between amino acids 2419 and 2420, by a new serine protease, contg. the sequence of AAR68621. The proteinase is purified as a fused product with the dithyrotolate reductase protein by using a methotrexate column. It can be used for the development of an inhibitor for HCV proteinase. (Updated on 16-Oct-2003 to standardise OS field)
XX	
SQ	Sequence 3010 AA;
Query Match	95.9%; Score 1705; DB 2; Length 3010;
Best Local Similarity	94.4%; Pred. No. 2.5e-155;
Matches 322; Conservative 11; Mismatches 8; Indels 0; Gaps 0;	
Dn	1 VRGRDAILITLCVHPRLIPDITKLLAATGPMVLQAAGTTKPYFVRAGILRACMLV 60   866 VRGGRDALITLCAVHPELPIDITKLLPILIGPLMVLQAGITRPYFVRAGILRACMLV 925 
Oy	61 RKAAGHYVQMAFMKLAAITGYVDHLTPLLQDMAHAGLRDLAVAVEPIFSMEVKLIIT 120   926 RKVAGHYVQMAFMKLAAITGYVDHLTPLLQDMAHAGLRDLAVAVEPIFSMEVKLIIT 985 
Dn	121 WGAOTTAACGDIIISGLPVASRRGKEILLGPDSFGEGQRLLAPITAYSQOTRGILGCIT 180   986 WGAOTTAACGDIIISGLPVASRRGKEILLGPDSFGEGQRLLAPITAYSQOTRGILGCIT 1045 
Oy	181 SLTGSDKKQVGEVGVVSTATOSFLATCNGVCWTFPGAGSKTLAPGKPITOWNTNVD 240   1046 SLTGSDKKQVGEVGVVSTATOSFLATCNGVCWTFVHGASSKTLAGKGIITQMYTNVD 1105 
Dn	241 QDLVGMOAPPGARSMTPTCTGSSSDLIVLTRADVI PVRRDDSRGSLSPRPVSYLKSS 300   1106 QDLVGMPAPPARSMTPTCTGSSSDLIVLTRADVI PVRRDDSRGSLSPRPISYLKSS 1165 
Oy	301 GGPLCPSGHAVGFRAVACTRGVAKAVDFIPVESMETMR 341   1166 GGPLCPSGHAVGFRAVACTRGVAKAVDFIPVESMETMR 1206 
Dn	
RESULT 8	
AAR68864	ID AAR68864 standard; protein; 3010 AA.
XX	
AC	AAR68864;
XX	
DT	06-DEC-1995 (first entry)
XX	
DE	Hepatitis C virus RNA helicase.
XX	
KW	Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
XX	Baculovirus; recombinant production.
XX	
OS	Hepatitis C virus.
XX	
FH	Key
FT	Region
FT	Location/Qualifiers
FT	196..198
FT	/label= N-linked glycosylation site
FT	209..211
FT	/label= N-linked glycosylation site
FT	234..236
FT	/label= N-linked glycosylation site
FT	250..252
FT	/label= N-linked glycosylation site
FT	305..307
FT	/label= N-linked glycosylation site

Query Match	95.8%	Score 1704	DB 2	Length 3010
Best Local Similarity	94.4%	Pred. No. 3.2e-155		
Matches 322	Conservative 11	Mismatches 8	Indels 0	Gaps 0
QY	1	VRGGRDAIILITCAVHEPILFDITKLLALIFGPIMLVQAGITKVPYFVARAGLIRACMLV	60	
DB	866	VRGRDAIILITCAVHEPILFDITKLLALIFGPIMLVQAGITKVPYFVARAGLIRACMLV	925	
QY	61	RKAAGHYVQMAFKKALITGTYVDHLTPLOQMAHAGLRDLAVAVEPVIFSDMEVKIIT	120	
DB	926	RKVAAGHYVQMAFKKALITGTYVDHLTPLOQMAHAGLRDLAVAVEPVIFSDMEVKIIT	985	
QY	121	WGADTTACGGIITGLPVSARRGRIILGPDADNFGGGMRIILAPITAYSOOTRGLGCIIT	180	
DB	986	WGADTTACGGIITGLPVSARRGRIILGPDADNFGGGMRIILAPITAYSOOTRGLGCIIT	1045	

QY 181 SLTGRDNQVGEVGVSTATQSFATCNVGCMTVPHGAGSKTLAAGPKPITQMTNTVD 240  
 DB 1046 SLTGRDNQVGEVGVSTATQSFATCNVGCMTVPHGAGSKTLAAGPKPITQMTNTVD 1105  
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVPVRRGDSRGSLLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVPVRRGDSRGSLLSPRPVSYLKSS 1165  
 QY 301 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 DB 1166 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206  
 RESULT 9  
 ABG30601 standard; protein; 2201 AA.  
 AC ABG30601;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT MISC-difference 882 /label= Arg, Lys  
 FT MISC-difference 2183 /note= "Wild type Met substituted by Thr"  
 FT  
 XX  
 PN WO200252015-A2.  
 PD 04-JUL-2002.  
 PF 20-DEC-2001; 2001WO-CA001843.  
 PR 22-DEC-2000; 2000US-0257857P.  
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj G, Pause A;  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apkx12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in

CC the claims of the invention  
 XX Sequence 2201 AA;  
 SQ  
 Query Match 95.7%; Score 1702; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;  
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 VAGSDAIIILTCANHPHLEIPDITKLLAIAGPLMVAQAGTKYKPYFPAAGILRACMLV 60  
 DB 57 VAGSDAIIILTCANHPHLEIPDITKLLAIAGPLMVAQAGTKYKPYFPAAGILRACMLV 116  
 QY 61 RRAAGHYVQMAFMKLAALTGYVVDHPLPLQDMAHAGLRDLAVAVEPVSDEMKIIT 120  
 DB 117 RYVAGHYVQMAFMKLAALTGYVVDHPLPLQDMAHAGLRDLAVAVEPVSDEMKIIT 176  
 QY 121 WGADTPACGDIISGLPVGARRGREIILGPADNFBGQWRLLAPITAYGQOTRGLGCIIT 180  
 DB 177 WGADTPACGDIISGLPVGARRGREIILGPADNFBGQWRLLAPITAYGQOTRGLGCIIT 236  
 QY 181 SLTGRDNQVGEVGVSTATQSFATCNVGCMTVPHGAGSKTLAAGPKPITQMTNTVD 240  
 DB 237 SLTGRDNQVGEVGVSTATQSFATCNVGCMTVPHGAGSKTLAAGPKPITQMTNTVD 296  
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVPVRRGDSRGSLLSPRPVSYLKSS 300  
 DB 297 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVPVRRGDSRGSLLSPRPVSYLKSS 356  
 QY 301 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 DB 357 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397  
 RESULT 10  
 ABG30591 standard; protein; 2201 AA.  
 ID ABG30591  
 XX  
 AC ABG30591;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT MISC-difference 751 /note= "Wild type Ser substituted by Gly"  
 FT MISC-difference 882 /label= Arg, Lys  
 FT  
 XX  
 PN WO200252015-A2.  
 PD 04-JUL-2002.  
 PF 20-DEC-2001; 2001WO-CA001843.  
 PR 22-DEC-2000; 2000US-0257857P.  
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj G, Pause A;  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX

PS Claim 3; Page; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APOK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

XX Sequence 2201 AA:

Query Match 95.7%; Score 1702; DB 5; Length 2201;

Best Local Similarity 94.1%; Pred. No. 3.2e-155;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIIILTCVHPELFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 60  
 DB 57 VRGGRDAVILITCAIHPELFTITKLLAILGPIMLVQAGITKVPYFVRAHGLIRACMLV 116  
 QY 61 RKAGAGHYVQMAFPMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIT 120  
 DB 117 RKVAGGHYVQMAFMKLAALGTYYVDHLTPLRMAHAGRLDAVAVEPIFSDMEVKIT 176  
 QY 121 WGAADTAACGDIISGLPVARSREIILGPADNFGGQWRLAPITAYVSOQTRGLGCIIT 180  
 DB 177 WGAADTAACGDIISGLPVARSREIILGPADNFGGQWRLAPITAYVSOQTRGLGCIIT 236  
 QY 181 SLTGRDNQVGEVQVSTATQSFATCVCVGTVEHGAQSKTLAAGPKGITOMYTND 240  
 DB 237 SLTGRDNQVGEVQVSTATQSFATCVCVGTVEHGAQSKTLAAGPKGITOMYTND 296  
 QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 300  
 DB 297 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 356  
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTKR 341  
 DB 357 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTKR 397

RESULT 11

ABG30600 standard; protein; 2201 AA.

XX ABG30600;

XX 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutuin.

OS Hepatitis C virus.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Msc-difference 882 /label= Arg, Lys

FT Msc-difference 1357 /note= "Wild type Pro substituted by Leu"

XX WO200252015-A2.

XX 04-JUL-2002.

XX 20-DEC-2001; 2001WO-CAN001843.

XX 22-DEC-2000; 2000US-0257857P.

XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;

XX WPI; 2002-575382/61.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.

PS Claim 3; Page; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon APOK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention

XX Sequence 2201 AA:

Query Match 95.7%; Score 1702; DB 5; Length 2201;

Best Local Similarity 94.1%; Pred. No. 3.2e-155;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIIILTCVHPELFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 60  
 DB 57 VRGGRDAVILITCAIHPELFTITKLLAILGPIMLVQAGITKVPYFVRAHGLIRACMLV 116  
 QY 61 RKAGAGHYVQMAFPMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPIFSDMEVKIT 120  
 DB 117 RKVAGGHYVQMAFMKLAALGTYYVDHLTPLRMAHAGRLDAVAVEPIFSDMEVKIT 176  
 QY 121 WGAADTAACGDIISGLPVARSREIILGPADNFGGQWRLAPITAYVSOQTRGLGCIIT 180  
 DB 177 WGAADTAACGDIISGLPVARSREIILGPADNFGGQWRLAPITAYVSOQTRGLGCIIT 236  
 QY 181 SLTGRDNQVGEVQVSTATQSFATCVCVGTVEHGAQSKTLAAGPKGITOMYTND 240  
 DB 237 SLTGRDNQVGEVQVSTATQSFATCVCVGTVEHGAQSKTLAAGPKGITOMYTND 296  
 QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 300  
 DB 297 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 356  
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTKR 341  
 DB 357 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTKR 397

RESULT 12

ABG30581

ID ABG30581 standard; protein; 2201 AA.  
 XX  
 AC ABG30581;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FN W0200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 DR N-PSDB; ABK85573.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Disclosure; Page 49-58; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon Apkx12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 CC  
 XX  
 SQ Sequence 2201 AA;  
 XX  
 Query Match 95.7%; Score 1702; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155; Indels 0; Gaps 0;  
 Matches 321; Conservative 11; Mismatches 9;  
 QY 1 VRGGRDAIILTCVAMELIFDITKLLAIFGPMVLQAGITKVPYFVRAGLIRACMLV 60  
 DB 57 VRGGRDAVILITCAIHPELIFPTIKLLAIFGPMVLQAGITKVPYFVRAGLIRACMLV 116  
 QY 61 RKAAGHYVQMAFMKLAALGTYYDHLTPQDMAHGLDIAVAVPVEFMEYKILIT 120  
 DB 117 RKAAGHYVQMAFMKLAALGTYYDHLTPQDMAHGLDIAVAVPVEFMEYKILIT 176  
 QY 121 WGAATAACGDIISGLPYASRRGRILLGPADNFEQGMRLAPITAYSOOTRGLGCIIT 180  
 DB 177 WGAATAACGDIISGLPYASRRGRILLGPADNFEQGMRLAPITAYSOOTRGLGCIIT 236  
 QY 181 SLTGRDKNOVGEVQVSTATQSLATCVNGVCTVHGAGSKTLAPKPKPIQMYTNVD 240  
 DB 237 SLTGRDKNOVGEVQVSTATQSLATCVNGVCTVHGAGSKTLAPKPKPIQMYTNVD 296  
 QY 241 QDLVGQAPPGARSMPTCTCGSSDLYIVTRHADVIPIVRRRGRSGSLSPRPVSYLKSS 300

DB 297 QDLVGQAPPGARSMPTCTCGSSDLYIVTRHADVIPIVRRRGRSGSLSPRPVSYLKSS 356  
 QY 301 GGPFLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 DB 357 GGPFLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397  
 RESULT 13  
 ID ABG30593  
 XX ABG30593 standard; protein; 2201 AA.  
 AC ABG30593;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.  
 XX  
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutetn.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FN Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882 /label= Arg, Lys  
 FT Misc-difference 892 /note= "wild type Leu substituted by Phe"  
 FT  
 PN W0200252015-A2.  
 PD 04-JUL-2002.  
 PF 20-DEC-2001; 2001WO-CA001843.  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC C virus replicon Apkx12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 CC  
 XX  
 SQ Sequence 2201 AA;  
 XX  
 Query Match 95.7%; Score 1702; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILLTCVAVHPELIFDTIKLLAIFGLPMLVQAGITKYPYFVRAQGLIRACMLV 60  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 57 VRGGRDAVILLTCVAVHPELIFDTIKLLAIFGLPMLVQAGITKYPYFVRAQGLIRACMLV 116

QY 61 RKAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGLRDLAVAEVPYFSDMEVKIIT 120  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 117 RKAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGLRDLAVAEVPYFSDMEVKIIT 176

QY 121 WGAADTAACGDIISGLPVASARGREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 180  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 177 WGAADTAACGDIISGLPVASARGREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 236

QY 181 SLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVTHGASGKTLAEPKGPITOMYTND 240  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 237 SLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVTHGASGKTLAEPKGPITOMYTND 296

QY 241 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRSDSGSLSPRPVSYLKSS 300  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 297 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRSDSGSLSPRPVSYLKSS 356

QY 301 GGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 341  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 357 GGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 397

RESULT 14  
 ABG30582  
 ID ABG30582 standard; protein; 2201 AA.

XX ABG30582;  
 AC  
 XX  
 DT 21-OCT-2002 (first entry)

XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.  
 DE  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.  
 XX  
 OS Hepatitis C virus.

XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /note= "Wild type Lys substituted by Lys or Arg"  
 FT Misc-difference 1233  
 FT /note= "Wild type Gly substituted by Cys"

XX W0200252015-A2.  
 XX  
 XX 04-JUL-2002.  
 PD  
 XX 20-DEC-2001; 2001WO-CA001843.  
 XX  
 XX 22-DEC-2000; 2000US-0257857P.  
 PR  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 XX Kukulj G, Pause A;  
 PI  
 XX WPI: 2002-575382/61.  
 DR N-FSDS; ABK88574.  
 XX  
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 XX Disclosure; Page 59-69; 140pp; English.  
 PS  
 XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-

CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

SO Sequence 2201 AA:

Query Match 95.7%; Score 1702; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;  
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILLTCVAVHPELIFDTIKLLAIFGLPMLVQAGITKYPYFVRAQGLIRACMLV 60  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 57 VRGGRDAVILLTCVAVHPELIFDTIKLLAIFGLPMLVQAGITKYPYFVRAQGLIRACMLV 116

QY 61 RKAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGLRDLAVAEVPYFSDMEVKIIT 120  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 117 RKAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGLRDLAVAEVPYFSDMEVKIIT 176

QY 121 WGAADTAACGDIISGLPVASARGREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 180  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 177 WGAADTAACGDIISGLPVASARGREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 236

QY 181 SLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVTHGASGKTLAEPKGPITOMYTND 240  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 237 SLTGRDKNOVEGEVQVSTATQSFATCNGVCMTVTHGASGKTLAEPKGPITOMYTND 296

QY 241 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRSDSGSLSPRPVSYLKSS 300  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 297 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRSDSGSLSPRPVSYLKSS 356

QY 301 GGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 341  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this  
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 357 GGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 397

RESULT 15  
 ABG30580  
 ID ABG30580 standard; protein; 2201 AA.

XX ABG30580;  
 AC  
 XX  
 DT 21-OCT-2002 (first entry)

XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.  
 DE  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 XX  
 OS Hepatitis C virus.

XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /note= "Encoded by ARG"

XX W0200252015-A2.  
 XX  
 XX 04-JUL-2002.  
 PD  
 XX 20-DEC-2001; 2001WO-CA001843.  
 XX  
 XX 22-DEC-2000; 2000US-0257857P.  
 PR  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 XX The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide  
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-



PI Kukulj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Disclosure; Page 69-74; 140pp; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)/C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 95.7%; Score 1702; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;  
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 VRGGRDAIILLTCVAPHELIPDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60  
 Db 57 VRGGRDAVILLTCAIHPELIFTIKLILGPMVLQAGITKVPYFVRAHGLIRACMLV 116  
 QY 61 RKAAGHYVQMAEFKLAALGTYYDHLTPLODWAHAGLRDLAAVEPVI FSDMEVKIIT 120  
 Db 117 RKAAGHYVQMAELKLAALGTYYDHLTPLODWAHAGLRDLAAVEPVI FSDMEVKIIT 176  
 QY 121 WGAADTAACGDIISGLPVASARREILGPADNPEGQGMRLAPITAYSQOTRGILGCIIT 180  
 Db 177 WGAADTAACGDIISGLPVASARREIHLGPADNPEGQGMRLAPITAYSQOTRGILGCIIT 236  
 QY 181 SLTGRDNQVGEVQVYVSTATQSLATCANGVCWTVFHAGSKTLAAGPKPITQMTNV 240  
 Db 237 SLTGRDNQVGEVQVYVSTATQSLATCANGVCWTVFHAGSKTLAAGPKPITQMTNV 296  
 QY 241 QDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300  
 Db 297 QDLVGMQAPPGARSLTCTGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 356  
 QY 301 GGPILCPGSHAVGIFRAVCTRGVAKAVDPIPVESMETMR 341  
 Db 357 GGPILCPGSHAVGIFRAVCTRGVAKAVDPIPVESMETMR 397

Search completed: May 6, 2004, 09:30:48  
 Job time : 47.4159 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 11.1056 Seconds  
(without alignments)  
2953.573 Million cell updates/sec

Title: US-10-650-585-14

Sequence: 1 VRGGRDAIILTCVHPEL.....RGVAKAVDPFVDSMETMR 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 28366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1711	96.2	3010	1 A45573	genome polyprotein
2	1710	96.2	3010	1 GNMVUJ	genome polyprotein
3	1685	94.8	3010	1 GNMVUW	genome polyprotein
4	1665	93.6	3010	1 S18030	genome polyprotein
5	1649	92.7	3010	1 GNMVTC	genome polyprotein
6	1565	88.0	3011	1 GNMVU3	genome polyprotein
7	1560	87.7	3011	1 S40770	genome polyprotein
8	1555	87.5	3011	1 GNMVUW	genome polyprotein
9	1362	76.6	3014	1 JCS620	genome polyprotein
10	1269	71.4	3033	1 J01303	genome polyprotein
11	1262	71.0	3033	1 GNMVU8	genome polyprotein
12	408.5	23.0	3005	2 T08841	polyprotein - dour
13	342.5	19.3	2970	2 T08839	polyprotein - marm
14	102.5	5.8	660	2 VHMW32	structural protein
15	101	5.7	600	2 B46642	signal recognition
16	101	5.7	600	2 G87392	DNA-directed DNA p
17	100.5	5.7	353	2 G87392	conserved hypochet
18	99.5	5.6	692	2 H71426	hypothetical prote
19	98	5.5	399	2 AH3038	conserved hypochet
20	97.5	5.5	399	2 C98247	hypothetical 50.8K
21	95.5	5.4	1085	2 T03531	cohn protein homol
22	95.5	5.4	470	2 JC4098	tetracycline 6-hyd
23	93.5	5.3	1380	2 T18309	receptor-adenylate
24	93	5.2	7463	2 T36248	CDA peptide synthe
25	92.5	5.2	706	2 S33761	transferrin precu
26	92.5	5.2	716	2 G83612	hypothetical prote
27	92	5.2	659	2 B44212	structural protein
28	91.5	5.1	3414	1 GNMVU6	genome polyprotein
29	91	5.1	904	2 A84212	hypothetical prote

30	90.5	5.1	868	2 H81775	aceticate hydratase
31	90	5.1	2796	2 JC4743	fatty-acid synthase
32	89.5	5.0	961	2 AE0375	probable cation-tr
33	88.5	5.0	652	2 C70688	probable nitrate r
34	87.5	4.9	3069	2 H70656	fatty-acid synthase
35	87.5	4.9	3412	1 GNMVU8	genome polyprotein
36	86.5	4.9	347	2 S4167	maleate dehydrogena
37	86.5	4.9	470	1 N41VW8	exo-alpha-sialidase
38	86.5	4.9	707	2 D84154	cadmium-transporti
39	86	4.8	338	1 S34984	UDPglucose 4-epime
40	85.5	4.8	348	2 H70549	probable pdhs prot
41	85.5	4.8	3076	2 A87058	fatty acid synthase
42	85	4.8	470	1 N41V9	exo-alpha-sialidase
43	84.5	4.8	1057	2 I38171	hugl protein - hum
44	84	4.7	233	2 T35594	hypothetical prote
45	84	4.7	423	2 S43967	p54-ralpha stress-a

ALIGNMENTS

RESULT 1

A45573 genome polyprotein - hepatitis C virus (strain J1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (HCV) genome from a single Japanese carrier: se  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C/Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
R:tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, A  
Virus Res. 23, 39-53, 1992  
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: se  
A:Reference number: A45573; PMID:92295714; PMID:1318627  
A:Accession: A45573  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3010 <TAN>  
A:Cross-references: GB:DI1168; GB:DD1171; NID:G221612; PIDN:BAA01943.1; PID:G221613  
A:Experimental source: HCV-JT  
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrophobic; nucleotide binding; P-loop; polyprotein; serine  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <NE1>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1616-1682/Product: nonstructural protein NS4 #status predicted <NS4>  
F:1683-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match

Best Local Similarity 95.3%; Pred. No. 4.7e-136;  
Matches 325; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY	1	VGGGRDAIILTCVHPELIDITLLAIGPVLVLAAGITKVVYFPAQGLIRACMLV	60
DB	866	VGGGRDAIILTCVHPELIDITLLAIGPVLVLAAGITKVVYFPAQGLIRACMLV	925
QY	61	RRAGGVYQVAFMLAALGTIVYDHTPLQDMNAAGRLDAVAVESVDSMEVKIT	120
DB	926	RRAGGVYQVAFMLAALGTIVYDHTPLQDMNAAGRLDAVAVESVDSMEVKIT	985
QY	121	WADDTAACGDIISGLPVASRRREILGPADNFEQGRMLAPITAYQOIRGLGICIT	180
DB	986	WADDTAACGDIISGLPVASRRREILGPADNFEQGRMLAPITAYQOIRGLGICIT	1045
QY	181	SLTGRDKVQEGEVVSTATOSPLATCNGCMTVFAGAGSKTLAIGKPIPTQYTWVD	240

Db 1046 SLTGRDKNQVEGEVQVSTATQSFATLTVNGVCMVTFHGAAGSKTLAAGPKPITQMTYND 1105  
QY 241 QDLVGMQAPPGARSMTPCTCGSSDLYLTVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300  
Db 1106 QDLVGMQAPPGARSLTPCTCGSSDLYLTVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 1165  
QY 301 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
Db 1166 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

## RESULT 2

GNMWCT  
genome polypeptide - hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
C:Accession: A39253; P50086  
R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimoto, Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
A:Reference number: A39253; MUID:91088550; PMID:2175903  
A:Accession: A39253  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <KAT>  
A:Cross-references: GB:P50208; NID:9221610; PIDN:BA1423.1; PID:9221611  
R:Kato, N.; Okoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variation  
A:Reference number: P50085  
A:Accession: P50086  
A:Molecule type: genomic RNA  
A:Residues: 2650-2707 <KAT>  
A:Experimental source: Japanese isolate  
A:Comment: The cleavage sites of this polypeptide have not been determined.  
C:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <NEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1616-1862/Product: nonstructural protein NS4b #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 96.2%; Score 1710; DB 1; Length 3010;  
Best Local Similarity 94.7%; Pred. No. 5.7e-136; Indels 0; Gaps 0;  
Matches 323; Conservative 11; Mismatches 7;  
QY 1 VRGGRDAIILITCAVHPELIFDITKLALIFGLPMLVQAGITKVPYFRAQGLIRACMLV 60  
Db 866 VRGGRDAIILITCAVHPELIFDITKLALIFGLPMLVQAGITKVPYFRAQGLIRACMLV 925  
QY 61 RKAAGHYVQNAFMKLAALITGYVDHLTPLODMAHAGLRLDVAVEVPSDMEVKIIT 120  
Db 926 RKAAGHYVQNAFMKLAALITGYVDHLTPLODMAHAGLRLDVAVEVPSDMEVKIIT 985  
QY 121 WGAATTAACGDIISGLPVSARSGREIILGPADNFEQGRMLLAPITAVSQTRGLGCIIT 180  
Db 986 WGAATTAACGDIISGLPVSARSGREIILGPADNFEQGRMLLAPITAVSQTRGLGCIIT 1045  
QY 181 SLTGRDKNQVEGEVQVSTATQSFATLTVNGVCMVTFHGAAGSKTLAAGPKPITQMTYND 240  
Db 1046 SLTGRDKNQVEGEVQVSTATQSFATLTVNGVCMVTFHGAAGSKTLAAGPKPITQMTYND 1105  
QY 241 QDLVGMQAPPGARSMTPCTCGSSDLYLTVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300

Db 1106 QDLVGMQAPPGARSMTPCTCGSSDLYLTVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 1165  
QY 301 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
Db 1166 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

## RESULT 3

GNMWCT  
genome polypeptide - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (BC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40244  
R:Chen, P.-J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 188, 102-113, 1992  
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
A:Reference number: A40244; MUID:92230206; PMID:1314449  
A:Accession: A40244  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <CHE>  
A:Cross-references: GB:M4754  
C:Superfamily: hepatitis C virus genome polypeptide; glycoprotein; hydrolase; nonstructural  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <NEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1616-1862/Product: nonstructural protein NS4b #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 94.8%; Score 1685; DB 1; Length 3010;  
Best Local Similarity 92.7%; Pred. No. 7.5e-134; Indels 0; Gaps 0;  
Matches 316; Conservative 14; Mismatches 11;  
QY 1 VRGGRDAIILITCAVHPELIFDITKLALIFGLPMLVQAGITKVPYFRAQGLIRACMLV 60  
Db 866 VRGGRDAIILITCAVHPELIFDITKLALIFGLPMLVQAGITKVPYFRAQGLIRACMLV 925  
QY 61 RKAAGHYVQNAFMKLAALITGYVDHLTPLODMAHAGLRLDVAVEVPSDMEVKIIT 120  
Db 926 RKAAGHYVQNAFMKLAALITGYVDHLTPLODMAHAGLRLDVAVEVPSDMEVKIIT 985  
QY 121 WGAATTAACGDIISGLPVSARSGREIILGPADNFEQGRMLLAPITAVSQTRGLGCIIT 180  
Db 986 WGAATTAACGDIISGLPVSARSGREIILGPADNFEQGRMLLAPITAVSQTRGLGCIIT 1045  
QY 181 SLTGRDKNQVEGEVQVSTATQSFATLTVNGVCMVTFHGAAGSKTLAAGPKPITQMTYND 240  
Db 1046 SLTGRDKNQVEGEVQVSTATQSFATLTVNGVCMVTFHGAAGSKTLAAGPKPITQMTYND 1105  
QY 241 QDLVGMQAPPGARSMTPCTCGSSDLYLTVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300  
Db 1106 QDLVGMQAPPGARSLTPCTCGSSDLYLTVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 1165  
QY 301 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
Db 1166 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206  
RESULT 4  
S18030  
genome polypeptide - hepatitis C virus (isolate JKI)

N;Contains: capsid protein C, envelope protein M, hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
 C/Species: hepatitis C virus  
 A/Variety: isolate JKI  
 C/Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 23-Mar-2001  
 A/Accession: S18030; S33570; A48332; S18029  
 R/Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 Submitted to the EMBL Data Library, September 1991  
 A/Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
 A/Reference number: S18028  
 A/Accession: S18030  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <HON>  
 A/Cross-references: EMBL:X61596; NID:G959478; PIDN:CAA3793.1; PID:G59479  
 A/Experimental source: isolate JKI from an individual  
 R/Honda, M.; Kaneko, S.; Uoura, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 128, 163-169, 1993  
 A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A/Reference number: A48332; MUID:93119270; PMID:8380322  
 A/Accession: S33570  
 A/Molecule type: genomic RNA  
 A/Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
 A/Cross-references: EMBL:X61591  
 A/Note: this sequence is inconsistent with the nucleotide translation  
 A/Note: the authors translated the codon AGG for residue 43 as Pro, TCG for residue 320 as Trp, and TTC for residue 771 as Ser  
 A/Note: Sequence extracted from NCBI database (NCBI:121747, NCBI:P.121748)  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:116-191/Product: capsid protein C #status predicted <CPC>  
 F:112-389/Product: major envelope protein E #status predicted <EM>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 93.6%; Score 1665; DB 1; Length 3010;  
 Best Local Similarity 92.7%; Pred. No. 3.7e-132;  
 Matches 316; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

1 VGGRRDAIILTCVAHPELIFDITKLLAFGLMVLQAGITVFPFVAAGIIRACMIV 60  
 DB 866 VGGRRDAIILTCVAHSELIFDITKLLAFGLMVLQAGITVFPFVAAGIIRACMIV 925  
 QY 61 RKAAGHYVQMAFMKLAALTGTYYVDHLPLOMAHAGLDLVAVEPVIFSDMEVKIIT 120  
 DB 926 RKAAGHYVQMAFMKLAALTGTYYVDHLPLOMAHAGLDLVAVEPVIFSDMEVKIIT 985  
 QY 121 WGAADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOQTRGLLCIIT 180  
 DB 986 WGAADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOQTRGLLCIIT 1045  
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 240  
 DB 1046 SLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 1105  
 QY 241 QDVGWQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 300  
 DB 1106 QDVGWQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 1165  
 QY 301 GGLLCPGSHAVGIFRAAVCTRGVAAKADFIPIESMETTR 341  
 DB 1166 GGLLCPGSHAVGIFRAAVCTRGVAAKADFIPIESMETTR 1206

GNMVT  
 genome polyprotein - hepatitis C virus  
 N;Contains: capsid protein C, envelope protein M, hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
 C/Species: hepatitis C virus  
 C/Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text\_change 19-Jan-2001  
 A/Accession: A3465  
 R/Rakimzawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
 A/Title: Structure and organization of the hepatitis C virus genome isolated from human  
 A/Reference number: A3465; MUID:91140698; PMID:1847440  
 A/Accession: A3465  
 A/Molecule type: genomic RNA  
 A/Residues: 1-3010 <TAK>  
 A/Cross-references: EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G329771  
 C/Superfamily: hepatitis C virus genome polyprotein  
 C/Keywords: ATP; capsid protein C, envelope protein M; glycoprotein; hydrolase; nonstructural  
 F:116-191/Product: capsid protein C #status predicted <CPC>  
 F:112-389/Product: major envelope protein E #status predicted <EM>  
 F:192-389/Product: major envelope protein E #status predicted <EM>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 92.7%; Score 1649; DB 1; Length 3010;  
 Best Local Similarity 92.1%; Pred. No. 8.3e-131;  
 Matches 313; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

2 RGGRRDAIILTCVAHPELIFDITKLLAFGLMVLQAGITVFPFVAAGIIRACMIV 61  
 DB 867 RGGRRDAIILTCVAHPELIFDITKLLAFGLMVLQAGITVFPFVAAGIIRACMIV 926  
 QY 62 KKAAGHYVQMAFMKLAALTGTYYVDHLPLOMAHAGLDLVAVEPVIFSDMEVKIIT 121  
 DB 927 KKAAGHYVQMAFMKLAALTGTYYVDHLPLOMAHAGLDLVAVEPVIFSDMEVKIIT 986  
 QY 122 GADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOQTRGLLCIIT 181  
 DB 987 GADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOQTRGLLCIIT 1046  
 QY 182 LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 241  
 DB 1047 LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 1106  
 QY 242 DLVGWQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 301  
 DB 1107 DLVGWQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 1166  
 QY 302 GGLLCPGSHAVGIFRAAVCTRGVAAKADFIPIESMETTR 341  
 DB 1167 GGLLCPGSHAVGIFRAAVCTRGVAAKADFIPIESMETTR 1206

A:Molecule type: mRNA  
 A:Residues: 13011 <CHO>  
 A:Cross-references: GB:062221; NID:9329873; PIDN:AAA5676.1; PID:9329874  
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peuteher, U.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A>Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A:Reference number: PQ0393; PMID:9226871; PMID:1316939  
 A:Accession: PQ0403  
 A:Molecule type: genomic RNA  
 A:Residues: 1577-1633 <CH>  
 A:Cross-references: DDBJ:D10128  
 A:Experimental source: Isolates E-b16  
 A:Accession: PQ0404  
 A>Status: Preliminary  
 A:Molecule type: genomic RNA  
 A:Residues: 1577-1633 <CH2>  
 A:Experimental source: Isolates E-b17  
 C:Superfamily: hepatitis C virus genome polypeptide; glycoprotein; hydrolase; nonstructura  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: major envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1635/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,446,476,532,540,556,576,623,645,1213,1255,2041,2077,22  
 Query Match 88.0%; Score 1565; DB 1; Length 3011;  
 Best Local Similarity 84.2%; Pred. No. 1,1e-123;  
 Matches 287; Conservative 28; Mismatches 26; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLLAIFGLMLOAGITVPFVPAQGLIRACMLV 60  
 DB 866 VRGRDVAIILMCAVHPFLVDITKLLAIFGLMLOAGITVPFVPAQGLIRACMLV 925  
 QY 61 RKAAGHYVQMAFPMKALNTGYVYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 120  
 DB 926 RKAIGHYVQMAVITKGLNTGYVYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 985  
 QY 121 WGAADTAACGDIISGLPVASARGREIILGPADNFEQGGWRLAPITAYSCQTRGLGCIIT 180  
 DB 986 WGAADTAACGDIISGLPVASARGREIILGPADNFEQGGWRLAPITAYSCQTRGLGCIIT 1045  
 QY 161 SLTGRDNQVGEVQVISTATQSFATCVCNVCMTVFGAGSKTLGPKPIQMTYND 240  
 DB 1046 SLTGRDNQVGEVQVISTATQSFATCVCNVCMTVFGAGSKTLGPKPIQMTYND 1105  
 QY 241 QDLVGMQAPPGARSMTPTCTGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMTPTCTGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 1165  
 QY 301 GGPIICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETMR 341  
 DB 1166 GGPIICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 7  
 S40770 genome polypeptide - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: S40770; PC1285  
 R:Okamoto, H.  
 Submitted to the EMBL Data Library, March 1992  
 A:Reference number: S40770

A:Accession: S40770  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3011 <OKA>  
 A:Cross-references: EMBL:D10749; NID:9221586; PIDN:BA01582.1; PID:9221587  
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,  
 Jpn. J. Exp. Med. 60, 167-177, 1990  
 A>Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A:Reference number: PC1284; PMID:91013116; PMID:2170712  
 A:Accession: PC1285  
 A:Molecule type: genomic RNA  
 A:Residues: 1-513 <OK2>  
 A:Cross-references: GB:000831; NID:9221511; PIDN:BA00705.1; PID:9221512  
 A:Experimental source: Isolate HC-J1  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: major envelope protein M #status predicted <EMP>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 Query Match 87.7%; Score 1560; DB 1; Length 3011;  
 Best Local Similarity 83.9%; Pred. No. 2.9e-123;  
 Matches 286; Conservative 27; Mismatches 28; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLLAIFGLMLOAGITVPFVPAQGLIRACMLV 60  
 DB 866 VRGRDVAIILMCAVHPFLVDITKLLAIFGLMLOAGITVPFVPAQGLIRACMLV 925  
 QY 61 RKAAGHYVQMAFPMKALNTGYVYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 120  
 DB 926 RKAIGHYVQMAVITKGLNTGYVYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 985  
 QY 121 WGAADTAACGDIISGLPVASARGREIILGPADNFEQGGWRLAPITAYSCQTRGLGCIIT 180  
 DB 986 WGAADTAACGDIISGLPVASARGREIILGPADNFEQGGWRLAPITAYSCQTRGLGCIIT 1045  
 QY 161 SLTGRDNQVGEVQVISTATQSFATCVCNVCMTVFGAGSKTLGPKPIQMTYND 240  
 DB 1046 SLTGRDNQVGEVQVISTATQSFATCVCNVCMTVFGAGSKTLGPKPIQMTYND 1105  
 QY 241 QDLVGMQAPPGARSMTPTCTGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMTPTCTGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 1165  
 QY 301 GGPIICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETMR 341  
 DB 1166 GGPIICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 8  
 S40770 genome polypeptide - hepatitis C virus (strain H)  
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu  
 C:Species: hepatitis C virus  
 A>Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A36814; A41546  
 R:Rincausse, G.; Zebende, S.; Lee, D.H.; Sugtani, M.; Nasoff, M.; Prince, A.M.  
 Submitted to GenBank, July 1992  
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c  
 A:Reference number: A36814  
 A:Accession: A36814  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:9329737; PID:AAA5534.1; PID:9329738  
 R:Inchasp, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar  
 A:Reference number: A41546; MUID:92052256; PMID:1658800  
 A:Contents: annotation  
 A>Note: neither amino acid nor nucleotide sequence is given  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <BPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis C virus genome polyprotein  
 F:1320-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: DEXH motif  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:156,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23  
 Query Match 87.5%; Score 1555; DB 1; Length 3011;  
 Best Local Similarity 83.9%; Pred. No. 7,6e-123;  
 Matches 286; Conservative 29; Mismatches 26; Indels 0; Gaps 0;

QY 1 VGGGRDAIILITCAVHEPILFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 60  
 Db 866 VGGGRDAIILITCAVHEPILFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 925  
 QY 61 RKAAGHYVQVAFMKLAALGTYYDHLTPLODMAHAGLRDLAAVEPVFESDMEVKIT 120  
 Db 926 RKAAGHYVQVAFMKLAALGTYYDHLTPLODMAHAGLRDLAAVEPVFESDMEVKIT 985  
 QY 121 WADTAACGDIISGLPVASRRREILGPADNFEQGMRLAPITVYVSOOTRGLGCIIT 180  
 Db 986 WADTAACGDIISGLPVASRRREILGPADNFEQGMRLAPITVYVSOOTRGLGCIIT 1045  
 QY 181 SLTGRDNQVGEHVQVSTATOSFLATCVNGCMTVFHAGSKTLAPKPIQMTYVND 240  
 Db 1046 SLTGRDNQVGEHVQVSTATOSFLATCVNGCMTVFHAGSKTLAPKPIQMTYVND 1105  
 QY 241 QDIVWQAPPGARSMTPTCTCGSSDLYLTRADVLPVRRGDSKSLSPRVSYLKSS 300  
 Db 1106 QDIVWQAPPGARSMTPTCTCGSSDLYLTRADVLPVRRGDSKSLSPRVSYLKSS 1165  
 QY 301 GGPIICPSGHAVGIFRAAVCTRGVAKAVDFIVESMETMR 341  
 Db 1166 GGPIICPSGHAVGIFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 9  
 J05620  
 genome polyprotein - hepatitis C virus (isolate H7H1480)  
 N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: J05620  
 R:Chamberlain, R.M.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A:Title: The complete coding sequence of hepatitis C virus genome 5a, the predominant  
 A:Reference number: J05620; MUID:97365593; PMID:9223423  
 A:Accession: J05620  
 A:Molecule type: mRNA  
 A:Residues: 1-3014 <C>  
 A:Cross-references: GB:Y13184  
 A:Experimental source: genotype 5a, which predominates in South Africa  
 A>Note: the translation of the nucleotide sequence is not complete in this paper  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <BPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:384-408/Region: hypervariable #status predicted  
 F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1008-1616/Product: hepatitis C virus genome polyprotein  
 F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
 F:1313-1318/Region: DEXH motif  
 F:1317-1320/Region: DEXH motif  
 F:1617-1863/Product: nonstructural protein NS4 #status predicted <N4A>  
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted  
 Query Match 76.6%; Score 1362; DB 1; Length 3014;  
 Best Local Similarity 71.3%; Pred. No. 1.7e-106;  
 Matches 243; Conservative 49; Mismatches 49; Indels 0; Gaps 0;

QY 1 VGGGRDAIILITCAVHEPILFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 60  
 Db 867 VGGGRDAIILITCAVHEPILFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 926  
 QY 61 RKAAGHYVQVAFMKLAALGTYYDHLTPLODMAHAGLRDLAAVEPVFESDMEVKIT 120  
 Db 927 RKAAGHYVQVAFMKLAALGTYYDHLTPLODMAHAGLRDLAAVEPVFESDMEVKIT 986  
 QY 121 WADTAACGDIISGLPVASRRREILGPADNFEQGMRLAPITVYVSOOTRGLGCIIT 180  
 Db 987 WADTAACGDIISGLPVASRRREILGPADNFEQGMRLAPITVYVSOOTRGLGCIIT 1046  
 QY 181 SLTGRDNQVGEHVQVSTATOSFLATCVNGCMTVFHAGSKTLAPKPIQMTYVND 240  
 Db 1047 SLTGRDNQVGEHVQVSTATOSFLATCVNGCMTVFHAGSKTLAPKPIQMTYVND 1106  
 QY 241 QDIVWQAPPGARSMTPTCTCGSSDLYLTRADVLPVRRGDSKSLSPRVSYLKSS 300  
 Db 1107 QDIVWQAPPGARSMTPTCTCGSSDLYLTRADVLPVRRGDSKSLSPRVSYLKSS 1166  
 QY 301 GGPIICPSGHAVGIFRAAVCTRGVAKAVDFIVESMETMR 341  
 Db 1167 GGPIICPSGHAVGIFRAAVCTRGVAKAVDFIVESMETMR 1207

RESULT 10  
 J01303  
 genome polyprotein - hepatitis C virus (isolate HC-95)  
 N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C:Accession: J01303  
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kural, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum  
 A:Reference number: J01303; MUID:92044440; PMID:1658196  
 A:Accession: J01303  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3033 <OKA>  
 A:Cross-references: GB:D00944; NID:9221650; PID:BA00792.1; PID:9221651  
 A:Experimental source: isolate HC-95 from a Japanese individual  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transme  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <BPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEB>  
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1011-1619/Product: hepatitis C virus genome polyprotein  
 F:1316-1321/Region: nucleotide-binding motif B  
 F:1320-1333/Region: DEXH motif  
 F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4A>  
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,477,554,542,558,578,627,649,1091,1217,1259,2038,28  
 Query Match 71.4%; Score 1269; DB 1; Length 3033;  
 Best Local Similarity 67.2%; Pred. No. 1.3e-98;  
 Matches 225; Conservative 49; Mismatches 63; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 60  
 DB VRGRDGIILWAVAILFCGVVFDITKMLAVLGPAYLLKGLITVFPVFRHALLRMCTAV 929  
 QY 61 RKAGGYYVQMAFMKLAALGTYYVDHLPLOMAHAGLBDLAAVEPVFSDMEVKIT 120  
 DB 930 RHLAAGYYVQMLALGRMTGTIYDHLTPMSDMAAGLRDLAVAEPIIFSMEXKIV 989  
 QY 121 WGAADTAACGDIISGLPVARSRGREILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180  
 DB 990 WGAETVACGDIHLGLPVSARKGREVLGPADGYTSKMSKLAPITAYAOQTRGLGCIIV 1049  
 QY 181 SLTGRDNQVGEVQVYSTATQSFATCVNGVCMVTHGAGSKTLGPKPIQMTYND 240  
 DB 1050 SMTGRDXTEDQAGETQVLSVTQTFGLTISGVLMTVHGANKTLGSRQPVYQMTYSAE 1109  
 QY 241 ODLVQMOAPPGARSMTCTCGSSDLYLVTBHADVIPVRRGDSRGLSPRPVSYLKSS 300  
 DB 1110 GDLVQWSPSPGTSKLEPCTCGAVDLYLVTBNADYIPARRGRDKRGLSPRPLSTLKSS 1169  
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTR 341  
 DB 1170 GGPVLCRSHAVGVFRAAVCSRGVAKSIDFIPVETLIDIVTR 1210

## RESULT 11

genome polyprotein - hepatitis C virus (strain HC-U8)  
 N:Contains: capsid protein C; envelope protein M; hepatitis B (BC 3.4.21.98) (nonstructu  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A40250; PMID:92230232; PMID:1314459  
 R:Okamoto, H.; Kunita, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;  
 Virology 189, 331-341, 1992  
 A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep  
 A:Reference number: A40250; MUID:92230232; PMID:1314459  
 A:Accession: A40250  
 A:Residues: 1-3033 <OKA>  
 A:Molecule type: genomic RNA  
 A:Cross-references: GB:D10988; GB:D01221; NID:9221608; PIDN:BA01761.1; PID:9221609  
 R:Chan, S.W.; Womish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
 A:Reference number: P00393; MUID:92268871; PMID:1316939  
 A:Accession: P00393  
 A:Molecule type: genomic RNA  
 A:Residues: 2678-2754 <CHA>  
 A:Cross-references: DDBJ:D10134  
 A:Experimental source: Isolate E-bi2  
 R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno  
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
 A:Title: Distribution of plural HCV types in Japan.  
 A:Reference number: P00554; MUID:92068204; PMID:11720309  
 A:Accession: P00554  
 A:Molecule type: mRNA  
 A:Residues: 2678-2729 <KAT>  
 A:Cross-references: GB:D10562; GB:D90518; NID:9221523; PIDN:BA01418.1; PID:9221524  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
 P:115/Product: capsid protein C #status predicted <CPC>  
 P:115/Product: envelope protein M #status predicted <BPM>  
 P:192-389/Product: major envelope protein E #status predicted <ME>  
 P:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 P:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
 P:1011-1619/Product: hepatitis B #status predicted <NS3>  
 P:1234-1241/Region: nucleotide-binding motif A (P-loop)

F:1316-1321/Region: nucleotide-binding motif B  
 F:1320-1323/Region: DEXH motif  
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>  
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,231

Query Match 71.0%; Score 1262; DB 1; Length 3033;  
 Best Local Similarity 66.0%; Pred. No. 4.9e-98;  
 Matches 225; Conservative 54; Mismatches 62; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 60  
 DB VRGRDGIILWAVAILHPLVFEVTKMLALIGPAYLLKASLRIPYFVRAHALLRVCTLV 929  
 QY 61 RKAGGYYVQMAFMKLAALGTYYVDHLPLOMAHAGLBDLAAVEPVFSDMEVKIT 120  
 DB 930 RHLAAGYYVQMLALGRMTGTIYDHLTPMSDMAAGLRDLAVAEPIIFSMEXKIV 989  
 QY 121 WGAADTAACGDIISGLPVARSRGREILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180  
 DB 990 WGAETVACGDIHLGLPVSARKGREVLGPADGYTSKMSKLAPITAYAOQTRGLGCIIV 1049  
 QY 181 SLTGRDNQVGEVQVYSTATQSFATCVNGVCMVTHGAGSKTLGPKPIQMTYND 240  
 DB 1050 SMTGRDXTEDQAGETQVLSVTQTFGLTISGVLMTVHGANKTLGSRQPVYQMTYSAE 1109  
 QY 241 ODLVQMOAPPGARSMTCTCGSSDLYLVTBHADVIPVRRGDSRGLSPRPVSYLKSS 300  
 DB 1110 GDLVQWSPSPGTSKLEPCTCGAVDLYLVTBNADYIPARRGRDKRGLSPRPLSTLKSS 1169  
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTR 341  
 DB 1170 GGPVLCRSHAVGVFRAAVCSRGVAKSIDFIPVETLIDIVTR 1210

## RESULT 12

polyprotein - douroucouli hepatitis GB virus A  
 C:Species: douroucouli hepatitis GB virus A  
 C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C:Accession: T08841  
 R:Reiter, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A:Reference number: Z16486; MUID:98120818; PMID:9460920  
 A:Accession: T08841  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3005 <ERK>  
 A:Cross-references: EMBL:A023425; NID:92828599; PIDN:AA040502.1; PID:92828600  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: polyprotein

Query Match 23.0%; Score 408.5; DB 2; Length 3005;  
 Best Local Similarity 31.3%; Pred. No. 8.9e-26;  
 Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 2 RGRD-----AIIITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRAC 57  
 DB 813 RGRDWRVTAWVAAGIIFPREVVCSPAHTA-----LAALIDSIDYLETL-ILTPA 864  
 QY 58 MLVRK-----AGHYVQMAFMKLAALGTYYVDHLPLOMAHAGLBDLAAVEPV 109  
 DB 865 QPARARLLDSITFLGDADLTPAFVRLERGVITLFOHGOVSKAAAILDGLVLEPV 924  
 QY 110 IFSDEVKIITWADTAACGDIISGLPVARSRGREILG--PADNFEQGWRLAPITAY 167  
 DB 925 SVTAADCVIYRAAATLACGQVVELPVVARGBEVLGVPPSVRALPPGVPAPVVV- 983  
 QY 166 SQQTGLGCIITSLTGRDNQVGEVQVYSTATQSFATCVNGVCMVTHGAGSKTLG 227  
 DB 984 MORGCFPSVYKSMIGRDERHEGSIVLGLSTTRSGTCVNGVMTTFHGSNRIILAG 1043





```

QY 15 VHPRL-----FDITKLLATFGPLMTVQAGI-----TKVPFVFAQGLIRACMYR 61
Db 154 VHDVLVLMGGEVSELPFAKGFVILLAGLVGKTTVCATLACTYIKQG--KSCMLI- 210
QY 62 KAAAGHYVQMAFMKLAAL---TGIVYVDHLTPLO--DMAHAGLRDLAVAVEPVI FSDMEY 116
Db 211 --AGDVYRPAIDQLVLIGEQVGPVYTAGTDVKPADIAKQGLKEAK-----NNVDY 261
QY 117 KIITMGADTAACGDIISGLFVSARGRRL-----LGPADNFEQGGWRLAPITAYSQGT 171
Db 262 VIM---DTAGRLQIDKGMDELKDVKKFLNPTFVLLVVDAMTQ--EAAALVTFNVEI 315
QY 172 RGLGCIITSLTGRDKNQVEGEVQVVS 198
Db 316 -GITGALITKLDGDSRGGAALSVKEYS 341

```

Search completed: May 6, 2004, 09:37:20  
 Job time : 12.1056 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.119 Seconds

(without alignments)  
2494.160 Million cell updates/sec

Title: US-10-650-585-14

Perfect score: 1778  
Sequence: 1 VRGGRDAIILLTCVAPHELI.....RGVAKAVDFIPVSEMETTMR 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1711	96.2	3010	1 POLG_HCVUT
2	1710	96.2	3010	1 POLG_HCVUT
3	1685	94.8	3010	1 POLG_HCVUT
4	1649	92.7	3010	1 POLG_HCVUT
5	1565	88.0	3011	1 POLG_HCV1
6	1555	87.5	3011	1 POLG_HCV1
7	1269	71.4	3033	1 POLG_HCV6
8	1262	71.0	3033	1 POLG_HCV6
9	102.5	5.8	660	1 VST2_HEVPA
10	102.5	5.8	660	1 VST2_HEVPA
11	102.5	5.8	660	1 VST2_HEVPA
12	101	5.7	600	1 PRO2_MOUSE
13	93.5	5.3	1380	1 CYRA_LEIDO
14	93	5.2	434	1 TOLB_CHLIE
15	92.5	5.2	706	1 TRFE_HORSE
16	92	5.2	659	1 VST2_HEVME
17	91.5	5.1	660	1 VST2_HEVME
18	91.5	5.1	660	1 VST2_HEVME
19	91.5	5.1	3414	1 POLG_TBEVH
20	90.5	5.1	3414	1 POLG_TBEVH
21	89.5	5.0	961	1 ATCU_YERPE
22	87.5	4.9	3412	1 POLG_TBEVS
23	86.5	4.9	347	1 MDHM_EUCUG
24	86.5	4.9	470	1 NRAM_IAMM
25	85	4.8	338	1 GALE_NETCO
26	86	4.8	433	1 DCUA_MOLSU
27	85	4.8	470	1 NRAM_IATRA
28	85	4.8	730	1 HELS_METVA
29	85	4.8	854	1 PWP2_SCHPO
30	85	4.8	3313	1 CLR3_RAT
31	84.5	4.8	1705	1 PTPV_MOUSE
32	84	4.7	309	1 UCP2_RAT
33	84	4.7	339	1 GPDA_CORFE

34	84	4.7	423	1 MK09_MOUSE	Q9wtu6 mus musculu
35	84	4.7	470	1 NRAM_IARUE	P31510 Influenza a
36	84	4.7	503	1 YAO2_MYCTU	O05586 mycobacteri
37	83.5	4.7	538	1 DAC_ACTSP	P39045 actinomadar
38	83.5	4.7	1399	1 PPOC_PSEAR	Q9hw9 pseudomonas
39	83	4.7	341	1 MDHM_BRANA	Q43744 brassica na
40	82.5	4.6	453	1 NRAM_TAVIL	P03470 Influenza a
41	82.5	4.6	711	1 HPPA_RHIME	Q8vz23 rhizodium m
42	82	4.6	309	1 UCP2_MOUSE	P70406 mus musculu
43	82	4.6	384	1 MK08_BRARE	Q9gd9s brachydanio
44	82	4.6	403	1 PGK_CHMU	O9pln4 chlamydia m
45	82	4.6	425	1 VIBR_RAT	P48974 rattus norv

## ALIGNMENTS

RESULT 1  
POLG\_HCVUT STANDARD; PRT; 3010 AA.  
AC Q00269;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate HC-JT) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepatitisin.  
NCBI\_TaxID=31642;  
OX NCBI  
RN SEQUENCE FROM N.A.  
RX MEDLINE=92295714; PubMed=1318627;  
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
RT Molecular cloning of hepatitis C virus genome from a single Japanese  
RT carrier: sequence variation within the same individual and among  
RT infected individuals.";  
RL Virus Res. 23:39-53(1992).  
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA) (N).  
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein C and RNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: D11168; AAA01943.1; -;  
CC PIR: A45573; A45573.  
CC MEROPS: S29.001; -;  
CC MEROPS: U39.001; -;  
CC Interpro: IPR009003; Cys\_ser\_trypsin.  
CC Interpro: IPR001410; DEAD.  
CC Interpro: IPR002522; HCV\_capsid.



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, D90208; BAA14233.1; -  
 CC PIR, A39253; GNVVCT.  
 CC HSSP, P26663; LXP.  
 CC MEROPS, S29.001; -  
 DR INTERPRO, IPR009033; Cys\_Ser\_trypsin.  
 DR INTERPRO, IPR001410; DEAD.  
 DR INTERPRO, IPR002522; HCV\_capsid.  
 DR INTERPRO, IPR002521; HCV\_core.  
 DR INTERPRO, IPR002519; HCV\_env.  
 DR INTERPRO, IPR002531; HCV\_NS1.  
 DR INTERPRO, IPR002518; HCV\_NS2.  
 DR INTERPRO, IPR000745; HCV\_NS4.  
 DR INTERPRO, IPR001490; HCV\_NS4b.  
 DR INTERPRO, IPR002868; HCV\_NS5a.  
 DR INTERPRO, IPR002166; HCV\_RdRp.  
 DR INTERPRO, IPR004109; Peptidase\_C29.  
 DR INTERPRO, IPR007095; RNA\_pol\_DS\_PS.  
 DR INTERPRO, IPR007094; RNA\_pol\_PSVLr.  
 DR Pfam, PF01543; HCV\_capsid; 1.  
 DR Pfam, PF01542; HCV\_core; 1.  
 DR Pfam, PF01539; HCV\_env; 1.  
 DR Pfam, PF01560; HCV\_NS1; 1.  
 DR Pfam, PF01538; HCV\_NS2; 1.  
 DR Pfam, PF02907; HCV\_NS3; 1.  
 DR Pfam, PF01006; HCV\_NS4a; 1.  
 DR Pfam, PF01001; HCV\_NS4b; 1.  
 DR Pfam, PF01506; HCV\_NS5a; 1.  
 DR Pfam, PF00271; helicase\_C; 1.  
 DR Pfam, PF00998; Viral\_RdRp; 1.  
 DR Pfam, PF0186062; HCV\_NS1; 1.  
 DR SMART, SM00487; DEXDC; 1.  
 KM Polypeptide, Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.  
 FT INIT\_MET 1  
 FT CHAIN 1  
 FT CHAIN 115  
 FT CHAIN 116  
 FT CHAIN 191  
 FT CHAIN 192  
 FT CHAIN 383  
 FT CHAIN 729  
 FT CHAIN 730  
 FT CHAIN 1006  
 FT CHAIN 1015  
 FT CHAIN 1616  
 FT CHAIN 1862  
 FT CHAIN 1863  
 FT CHAIN 2013  
 FT CHAIN 2014  
 FT TRANSMEM 347  
 FT ACT\_SITE 1083  
 FT ACT\_SITE 1107  
 FT ACT\_SITE 1165  
 FT ACT\_SITE 1165  
 FT NP\_BIND 1230  
 FT NP\_BIND 1237  
 FT NP\_BIND 1319  
 FT NP\_BIND 1319  
 FT CARBOHYD 196  
 FT CARBOHYD 209  
 FT CARBOHYD 209  
 FT CARBOHYD 234  
 FT CARBOHYD 250  
 FT CARBOHYD 305  
 FT CARBOHYD 305  
 FT CARBOHYD 417  
 FT CARBOHYD 417  
 FT CARBOHYD 423  
 FT CARBOHYD 430  
 FT CARBOHYD 430  
 FT CARBOHYD 448  
 FT CARBOHYD 532  
 FT CARBOHYD 556  
 FT CARBOHYD 576  
 FT CARBOHYD 623  
 FT CARBOHYD 623  
 FT CARBOHYD 645  
 FT CARBOHYD 645  
 FT CARBOHYD 2041  
 FT CARBOHYD 2041

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2768 2768 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3010 AA; 327017 KM; AA993794F46DB185 CRC64;  
 CC SEQUENCE  
 Query Match 96.2%; Score 1710; DB 1; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 1.7e-136;  
 Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 VRGGSDAIIITLCAVPELIPDITKLLAIFGPMVLVAGITKVPYFRAQGLRACMLV 60  
 DB 866 VRGGSDAIIITLCAVPELIPDITKLLAIFGPMVLVAGITKVPYFRAQGLRACMLV 925  
 QY 61 RRAAGHYVQNAFMKLAALGTGYVDHUTPLQDMWAGLRDLAVAEVIFSDMEVKIIT 120  
 DB 926 RRAAGHYVQNAFMKLAALGTGYVDHUTPLQDMWAGLRDLAVAEVIFSDMEVKIIT 985  
 QY 121 WADPTAACGDIISGLPVARFRETLLGPADPEQGRLLAPITAYSOOTRGLIGCIIT 180  
 DB 986 WADPTAACGDIISGLPVARFRETLLGPADPEQGRLLAPITAYSOOTRGLIGCIIT 1045  
 QY 181 SLTGKDKQVBEVGVVSTATQSFATCVNGVCTVFPAGSKTLGPKPIYQYITVD 240  
 DB 1046 SLTGKDKQVBEVGVVSTATQSFATCVNGVCTVFPAGSKTLGPKPIYQYITVD 1105  
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLYVTRHADVTPVRRGDSRGLSPRVSLKSS 300  
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLYVTRHADVTPVRRGDSRGLSPRVSLKSS 1165  
 QY 301 GGELICPSGHAIVGIFRAVCTRGVAKAVDPFVSEMETTMR 341  
 DB 1166 GGELICPSGHAIVGIFRAVCTRGVAKAVDPFVSEMETTMR 1206  
 RESULT 3  
 POLG\_HCVTM STANDARD; PRT; 3010 AA.  
 AC P29846;  
 DT 01-APR-1993 (rel. 25, Created)  
 DT 01-APR-1993 (rel. 25, Last sequence update)  
 DT 10-OCT-2003 (rel. 42, Last annotation update)  
 DE Genome polypeptide [contains: Capsid protein C (Core protein) (P222);  
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2  
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=31645;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9230206; PubMed=1314449;  
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RT "The Taiwanese hepatitis C virus genome: sequence determination and  
 RT mapping the 5' terminal of viral genome and antigenomic RNA";  
 RL Virology 188:102-113(1992).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

Query Match	Best Local Similarity	Score	DB 1	Length	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB 400	DB 401	DB 402	DB 403	DB 404	DB 405	DB 406	DB 407	DB 408	DB 409	DB 410	DB 411	DB 412	DB 413	DB 414	DB 415	DB 416	DB
-------------	-----------------------	-------	------	--------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	----









	FT	TURND	1419	1420
FT	STRAND	1432	1436	
FT	TURND	1438	1439	
FT	STRAND	1450	1453	
FT	STRAND	1456	1463	
FT	STRAND	1471	1478	
FT	STRAND	1480	1480	
FT	HELIX	1481	1488	
FT	TURND	1489	1490	
FT	STRAND	1497	1501	
FT	STRAND	1507	1507	
FT	STRAND	1511	1511	
FT	HELIX	1514	1527	
FT	HELIX	1532	1544	
FT	STRAND	1550	1550	
FT	HELIX	1555	1564	
FT	HELIX	1570	1578	
FT	TURND	1579	1580	
FT	HELIX	1584	1597	
FT	TURND	1598	1598	
FT	HELIX	1606	1611	
FT	TURND	1614	1618	
FT	STRAND	1622	1623	
FT	STRAND	1627	1627	
FT	STRAND	1635	1636	
FT	HELIX	1640	1652	
SC	SEQUENCE	3011 AA;	327142 MW;	772CBB29CCD94753 CRC64;
Query Match		87.5%;	Score 1555;	DB 1;
Best Local Similarity		83.9%;	Pred. No. 2,3e-123;	Length 3011;
Matches	286;	Conservative	29;	Mismatches 26;
			Indels	0;
			Gaps	0;
QY	1	VRGRDAIILITCAVHPELLFDITKLLIAIFGPIMLVIOAGITKVPYFVRAGGLIRACMLV	60	
DB	866	VRGRDRDAVILITCVHFAVFDITKLLIAIFGPIMLVIOAGITKVPYFVRAGGLIRACMLV	925	
QY	61	PKAGGHVQWAEFKLALNGTYYVDLTPIDQAHAGLRLDAVAVEPVIFSDMEVKIIT	120	
QY	926	RKINGHVVQALIKGLTGTCVYNHAPURMAHNGRLDAVAVEPVIFSRMETKLIIT	985	
DB	926	RKINGHVVQALIKGLTGTCVYNHAPURMAHNGRLDAVAVEPVIFSRMETKLIIT	985	
QY	121	WGADTTAACGDIISGLPYSAARRGREILLGPADNFEQGMRLAPITAYSQOTRGLGCIIT	180	
DB	986	WGADTTAACGDIISGLPYSAARRGREILLGPADNFEQGMRLAPITAYSQOTRGLGCIIT	1045	
QY	181	SLTRDKKQVEGEVQVYSTATQSLATCANGVCMTVTHGAGSKTLAAPKGIITQMTNNVD	240	
DB	1046	SLTRDKKQVEGEVQVYSTATQSLATCANGVCMTVTHGAGSKTLAAPKGIITQMTNNVD	1105	
QY	241	QDLVGMQAPPARSMTPTCTGSSDLYIVTVTRHADVIPRRRGDSRGLSLSPRPVSYLKGSS	300	
DB	1106	QDLVGMQAPPARSMTPTCTGSSDLYIVTVTRHADVIPRRRGDSRGLSLSPRPVSYLKGSS	1165	
QY	301	GGPLLCFSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTNR	341	
DB	1166	GGPLLCFSGHAGVIFRAAVCTRGVAKAVDFIPVENLETITNR	1206	
RESULT 7				
POLG_HCVJ6		STANDARD;	PRT;	3033 AA.
AC	P26660;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Envelope glycoprotein [Contains: Capsid protein C (Core protein) (P22);			
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2			
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)			
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepactivin)			
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein			
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein			
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OC	Hepatitis C virus [isolate HC-J6] (ACV).			
OC	Vitruves; ssRNA positive-strand vitruves, no DNA stage; Flaviviridae;			



(EC 3.4.21.98): Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
 CC NCBI\_TaxId=11115;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9230233; Pubmed=1314459;  
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Iizuka H., Tanaka T., Fukuda S., Tsuda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes";  
 RT Virology 188:333-341 (1992).  
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).  
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.1sb-sib.ch/announce/> or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 CC EMBL: D10988; BAA01761.1; -;  
 CC PIR: A40250; GNWVJ8.  
 CC HSSP: P27958; 1HE1.  
 CC MEROPS: S29.001; -;  
 CC MEROPS: U39.001; -;  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4A.  
 DR InterPro: IPR001490; HCV\_NS4B.  
 DR InterPro: IPR002868; HCV\_NS5A.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSTvir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4A; 1.  
 DR Pfam: PF01001; HCV\_NS4B; 1.  
 DR Pfam: PF01506; HCV\_NS5A; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR Pfam: PD186062; HCV\_NS1; 1.  
 DR SMART: SMO0487; DEXDC; 1.  
 DR Polypeptide: Glycoprotein: Transferase: RNA-directed RNA polymerase;  
 KM Core protein: Coat protein: Envelope protein; Helicase; ATP-binding;  
 KM Transmembrane; Nonstructural protein: Hydrilase; Serine protease;  
 FT INIT\_MOT 1  
 FT 1 REMOVED FROM CAPSID PROTEIN C BY THE

Query Match	Best Local Similarity	Score 1262; DB 1; Length 3033;
Query 1	VEGRDAILLLTCAVPELIFDITLTLAIGPLVLTGATKVPYFVAGLIRACMLV	60
DB 870	VRGGRDGIIMWAVILHPLVFEVTKMLAIIIGPAVILKASLRIIPYFRAHALRVCTLV	929
Query 61	KRAAGGHVQVAFMKLALITGVYVDHITPIQDMAHAGLPLAVAVEVFSDEVKIIT	120
DB 930	KHLGARYIOMLLITIGMTGYIYDHSPLSTWAAQGLRDLAIVEVFSDEVKIV	989
Query 121	WGADTAAGDIIISGLPVARSREILIPADNFBGQRLIAPITAYSOQTRGLGCIIT	180
DB 990	WGAETVACGDIILHGLPVARLREVLISLPAGSYNSKMKLAPITAYQOTRGLLGAIV	1049
Query 181	SLTGRDKNQVEGEVQVSTAFQSLATVNGVCVTFVFGAGSKTLAGKGIITOMTVND	240
DB 1050	SLTGRDKNQVEGEVQVSTAFQSLATVNGVCVTFVFGAGSKTLAGKGIITOMTVND	1109
Query 241	ODLVGMQAPPGARSMTPCTGSSDILYVTRADYVIVARRSDPSGSLSPVSYLKSS	300
DB 1110	GDVGMPSPPGKSLDPTCGAVDLYVTRADYVIVARRSDPSGSLSPVSYLKSS	1169
Query 301	GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR	341
DB 1170	GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR	1210

RESULT 9  
 ID VST2 HEVBU STANDARD; PRT; 660 AA.  
 AC P29326; 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2)  
 OS Hepatitis E virus (strain Burma) (HEV)  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 CC Hepatitis E-like viruses.  
 NCBI\_TaxID=31767;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92024067; PubMed=1926770;  
 RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,  
 RA Fry K.E., Reyes G.R.;  
 RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the  
 RT full-length viral genome."  
 RL Virology 185:120-131(1991).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M73218; AAA45736.1; -  
 CC PIR: C40778; VHWMH2.  
 CC InterPro: IPR004261; SP2.  
 CC InterPro: IPR008975; Viral\_cap\_coat.  
 CC Pfam: PF03014; SP2; 1.  
 CC Signal.  
 FT CHAIN 1 19 POTENTIAL.  
 FT SIGNAL 20 660 STRUCTURAL PROTEIN 2.  
 SQ SEQUENCE 660 AA; 70978 MW; 5832A0130CCCA61C CRC64;  
 Query Match 5.8%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 0.55;  
 Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;  
 QY 24 TKLL--AIFGLMVLQAG-----ITKPYFVR--AAGLRACMLYKAGGHVQMA 72  
 DB 151 TNLVLYAAPLSPLPLQDGTNTHIMATEASNAQRYRATIRYPLVFNAGVYALIS 210  
 QY 73 FMKLAALGTYYVDHLTPLODWAHAGLDLAVAEVPIFSDMEVKITWGADTAACGDI 132  
 DB 211 FWPQTTTPTSV-----DNMSITSDVRLVPGIASELVI----- 246  
 QY 133 SGLPVASARRREILIGPAD--NFGQGMRLAP--TVSQQTRGL-----GCIITSNG 184  
 DB 247 -----PSERLHYRNQGMRSVETSGVAEEATSGVLVCLIHGSLVNSYTN 290  
 QY 185 -----RDKNQVEGEVNVSTATQSF 205  
 DB 291 TPTGALGLDPALEFRNLTPGNTNTRVRSYSTARHRLRGADGTALITTTAATRM 350  
 QY 206 A---TCVNGV-----CMTVFH-----GAG-----SKTLAPKG--PIT 233  
 DB 351 KDLYFTSTNGVEIGRIALTLFNLADTLIGLPTELISAGGOLFYSRPVVSANGEPTV 410  
 QY 234 QMYTNVDQDLVGMQAPGARSMTPCTGSSDLVYV---TRHADVIPIVRRGDSRG--SLIS 289  
 DB 411 KLTYSVENA---QODKGIAPHDIDLGESRVVIOVDNQHEDRPPTSPAPSRFFSVLR 466  
 QY 290 PRPVSYLK-----GSSGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPV 333  
 DB 467 ANDVLMILSLTAEXDSTYSGSTGPVYV--SDSVTLVNVATGAQAVASIDMTKV 519

DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2)  
 OS Hepatitis E virus (strain Pakistan) (HEV)  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 CC Hepatitis E-like viruses.  
 NCBI\_TaxID=33774;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92115700; PubMed=1713127;  
 RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,  
 RA Malik I.A., Iqbal M., Purcell R.H.;  
 RT "Characterization of a prototype strain of hepatitis E virus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING  
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA  
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M80581; AAA45727.1; -  
 CC InterPro: IPR004261; SP2.  
 CC InterPro: IPR008975; Viral\_cap\_coat.  
 CC Pfam: PF03014; SP2; 1.  
 CC Signal.  
 FT CHAIN 1 22 BY SIMILARITY.  
 FT SIGNAL 23 660 STRUCTURAL PROTEIN 2.  
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC3C7B46FD3 CRC64;  
 Query Match 5.8%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 0.55;  
 Matches 82; Conservative 54; Mismatches 128; Indels 151; Gaps 20;  
 QY 24 TKLL--AIFGLMVLQAG-----ITKPYFVR--AAGLRACMLYKAGGHVQMA 72  
 DB 151 TNLVLYAAPLSPLPLQDGTNTHIMATEASNAQRYRATIRYPLVFNAGVYALIS 210  
 QY 73 FMKLAALGTYYVDHLTPLODWAHAGLDLAVAEVPIFSDMEVKITWGADTAACGDI 132  
 DB 211 FWPQTTTPTSV-----DNMSITSDVRLVPGIASELVI----- 246  
 QY 133 SGLPVASARRREILIGPAD--NFGQGMRL--APITAYSQ 169  
 DB 247 -----PSERLHYRNQGMRSVETSGVAEEATSGVLVCLIHGSLVNSYTN 290  
 QY 170 QT-RGLGCI-----ITSLGRDKNQ-----VEGEVNVSTATQSF 205  
 DB 291 TPTGALGLDPALEFRNLTPGNTNTRVRSYSTARHRLRGADGTALITTTAATRM 350  
 QY 206 A---TCVNGV-----CMTVFH-----GAG-----SKTLAPKG--PIT 233  
 DB 351 KDLYFTSTNGVEIGRIALTLFNLADTLIGLPTELISAGGOLFYSRPVVSANGEPTV 410  
 QY 234 QMYTNVDQDLVGMQAPGARSMTPCTGSSDLVYV---TRHADVIPIVRRGDSRG--SLIS 289  
 DB 411 KLTYSVENA---QODKGIAPHDIDLGESRVVIOVDNQHEDRPPTSPAPSRFFSVLR 466  
 QY 290 PRPVSYLK-----GSSGGLPCPSGHAVGIFRAAVCTRGVAKAVDFIPV 333  
 DB 467 ANDVLMILSLTAEXDSTYSGSTGPVYV--SDSVTLVNVATGAQAVASIDMTKV 519

RESULT 10  
 VST2\_HEVPA  
 ID VST2\_HEVPA  
 AC P33426;  
 STANDARD;  
 PRT;  
 660 AA.

RESULT 11  
 SRSC\_ARATH  
 ID SRSC\_ARATH  
 AC P37107; 082570;  
 STANDARD;  
 PRT;  
 564 AA.

DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Signal recognition particle 54 kDa protein, chloroplast precursor  
 DE (SRP54) (54 chloroplast protein) (54CP) (FFC).  
 GN PFC OR AT5G03940 OR PFC6\_150.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC Eucrotyd II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RX STRAIN=cv. Columbia;  
 RX MEDLINE=94012817; PubMed=8408079;  
 RA Franklin A.E., Hoffman N.E.;  
 RT "Characterization of a chloroplast homologue of the 54-kDa subunit of  
 the signal recognition particle";  
 RL J. Biol. Chem. 268:22175-22180(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;  
 RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,  
 encoding the 54 kDa subunit of chloroplast signal recognition  
 particle";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016721; PubMed=11130714;  
 RA Tabata S., Kaneo T., Nakamura Y., Kotani H., Kato T., Asanizu E.,  
 RA Miyajima N., Sasekubo S., Kimura T., Hosouchi T., Kawashima K.,  
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
 RA Nakazaki N., Natsu K., Okumura S., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yanada M., Yasuda M., Sato S., de la Bastide M.,  
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordun H., Cordes J., Haakensen B., Lamar E., Latreille P.,  
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,  
 RA Wagner-McPherson C., Wolian A., Yoakum M., Bell M., Dedila N.,  
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
 RA Mattiensen R., McComb W.R., Wilson R.K., Murphy G., Bancroft I.,  
 RA Volckert G., Wambut R., Duesternoeft A., Stiekema W., Pohl T.,  
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,  
 RA Langham S.-A., McCullagh B., Robben J., Grymptre B., Zimmermann W.,  
 RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,  
 RA van Steveren M., Dikse G., Rose M., Moolman P., Klein Lankhorst R.,  
 RA Weitzenecker T., Bothe G., Rose M., Haut J., Bernerstorff S., Hempel S.,  
 RA Feldpausch M., Lamberth S., Villarroel R., Gelsen J., Ardiles W.,  
 RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franse P.,  
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 thaliana";  
 RL Nature 408:823-826(2000).  
 RL -1- FUNCTION: May target chloroplast proteins to either the thylakoid  
 or envelope membranes.  
 CC -1- SUBCELLULAR LOCATION: Chloroplast stroma.  
 CC -1- TISSUE SPECIFICITY: Most abundant in green shoot tissue and  
 lower levels seen in the roots and etiolated buds.  
 CC -1- SIMILARITY: Belongs to the GTP-binding SRP family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, Z21970; CAA79981.1; -

DR EMBL; AF092168; AAC64139.1; -  
 DR EMBL; AL162873; CAB85514.1; -  
 DR PIR; S36637; S36637.  
 DR HSPD; 007347; IFFH.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR000897; SRP54.  
 DR InterPro; IPR004125; SRP54\_SBP.  
 DR InterPro; IPR004780; SRP\_SBP.  
 DR Pfam; PF00448; SRP54\_1.  
 DR Pfam; PF02881; SRP54\_N; 1.  
 DR Pfam; PF02978; SRP\_SBP; 1.  
 DR ProDom; PD000819; SRP54\_1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFam; TIGR00959; ffb; 1.  
 DR PROSITE; PS00300; SRP54\_1.  
 KW Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;  
 KW Transist peptide.  
 FT TRANSIT 1  
 FT CHAIN 76  
 FT DOMAIN 76 370  
 FT DOMAIN 371 564  
 FT NP\_BIND 183 190  
 FT NP\_BIND 265 269  
 FT NP\_BIND 323 326  
 FT CONFLICT 76 76  
 SQ SEQUENCE 564 AA; 61232 MW; 423F7285FB9063E4 CRC64;  
 Query Match 5.7%; Score 101; DB 1; Length 564;  
 Best Local Similarity 26.1%; Freq. No. 0.6; Indels 42; Gaps 11;  
 Matches 54; Conservative 37; Mismatches 74;  
 QY 15 VPEIET-----PDITLLAIGPLMVLQAGI-----TKVFFVPAQGLIRACMLVR 61  
 DB 154 VHDVLYKLKMGSEVSELOKSGPVTLLAGLGQVGTCTVCKLACVLYKKQ--KSCWLI- 210  
 QY 62 KAAGHYVQVAPFKIAL--TGYVYDHLPLQ--DVAHGLRDLAVAVPVFPMMEV 116  
 DB 211 --AGVYPPAIDVLVILGEGVGVVYAGTVKPADAKGLKLEAK-----NNVDV 261  
 QY 117 KITTGADTAACGDISGLPVSARGREIL-----LGPADNFEQGNLLAPITAVSQOT 171  
 DB 262 VIM---DTAGRLQIDKGMDELKDKFLNPTFVLVDMTQ--EAAVLVTFNVEI 315  
 QY 172 RGLGCIITSLTGRDNQGEVEQVYS 198  
 DB 316 -GITGAILTKLDGDSRGALSVKVS 341  
 RESULT 12  
 DP02 MOUSE STANDARD; PRT; 600 AA.  
 AC P33611;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).  
 GN POLA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 265-265 AND 394-403.  
 RX MEDLINE=93216788; PubMed=8463324;  
 RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,  
 RA Hanaoka F.;  
 RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA  
 polymerase alpha-primase complex and their gene expression during  
 cell proliferation and the cell cycle";  
 RL J. Biol. Chem. 265:8111-8122(1993).  
 CC -1- FUNCTION: May play an essential role at the early stage of  
 chromosomal DNA replication by coupling the polymerase

```
CC alpha/primase complex to the cellular replication machinery (By
CC similarity).
CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
CC (subunits A, B, C and D), which is assembled throughout the cell
CC cycle. The largest subunit (subunit A) has DNA polymerase
CC activity, the two smallest subunits (subunits C and D) have DNA
CC primase activity. Subunit B binds to subunit A.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC PHASE (BY SIMILARITY).
CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D13546; BAA02746.1; -.
CC PIR: B46642; B46642.
CC MGI: 99690; Polad.
CC InterPro: IPR007200; DNA_pol_alpha_B.
CC Pfam: PF04058; DNA_pol_alpha_B; 1. Phosphorylation.
CC KEGG: K01107; POLY-GLU.
CC FT DOMAIN 101 107
CC PRO/SER/THR-RICH (HYDROPHILIC).
CC SQ SEQUENCE 600 AA; 66267 MW; 79F94BEEF3FEEC CRC64;
CC
CC Query Match 5.7%; Score 101; DB 1; Length 600;
CC Best Local Similarity 24.8%; Pred. No. 0.65; 71; Indels 62; Gaps 12;
CC Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;
CC
CC 53 LIRACMVERKAGGHYOM-APMKLAALT-----GYVYDHL-----TPLDWA 95
CC 27 LAELCVLYRDEDMVSELIFCTSAAGTCLTVDILNSFEVYANKLSKAMSASKDSG 86
CC
CC QY 96 HAGRIDLAIVEPIVPSDMEYKITTWADTAACDI--ISGLP-----VSARRGRET 145
CC DB 87 HAGRIDLAIVEPIVPSDMEYKITTWADTAACDI--ISGLP-----VSARRGRET 144
CC
CC QY 146 LIGPADNFEQGGWLLPITAYSOQTRGILGCIITSLNGRDKXQVEVEQVSTANGSFL 205
CC DB 145 LLSBS-----FSPSATPSOK-----YTSSTR-----GEVTTGSAQ-- 178
CC
CC QY 206 ATCVNGVCMVTFHAGSKTL--AGPKPIPTOMYTNVDQDLVG 245
CC DB 179 -----GLWSGSGSGSVSLKVGDPPLTGSYVAMFQQLMG 215
CC
CC RESULT 13
CC CYAA LEIDO STANDARD; PRT; 1380 AA.
CC ID CYAA LEIDO STANDARD; PRT; 1380 AA.
CC AC Q27675;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Reciprocal-type adenylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-
CC lyase) (adenylate cyclase).
CC GN RAC-A.
CC OS Leishmania donovani.
CC OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
CC NX NCBI_TaxID=5661;
CC [1]
CC R2 SEQUENCE FROM N.A.
CC RC STRAIN=IS Sudanese;
CC RX MEDLINE=9534054; PubMed=7615561;
CC RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.,
CC RT "A family of putative receptor-adenylate cyclases from Leishmania
CC donovani.";
CC RL J. Biol. Chem. 270:17551-17558 (1995).
CC -1- FUNCTION: Could act as a receptor for a unknown ligand.
```

```
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote)
CC but not in the mammalian host stage of the parasite life cycle.
CC -1- SIMILARITY: Belongs to the adenylate cyclase class-3 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U17042; AAA7498.1; -.
CC PIR: T18309; T18309.
CC InterPro: IPR001054; G_cyclase.
CC Pfam: PF00211; guanylate_cyc; 1.
CC SMART: SM00044; CYCC; 1.
CC DR PROSITE: PS50125; GUANYLATE CYCLASES 2; 1.
CC KW Metal-binding; Magnesium;
CC Metal-binding; Magnesium.
CC FT DOMAIN 1 34
CC FT TRANSMEM 35 55
CC FT DOMAIN 56 891
CC FT TRANSMEM 892 912
CC FT DOMAIN 913 1380
CC FT METAL 938 938
CC FT METAL 981 981
CC FT CARBOHYD 422 422
CC FT CARBOHYD 478 478
CC FT CARBOHYD 497 497
CC FT CARBOHYD 567 567
CC SQ SEQUENCE 1380 AA; 151692 MW; 6B2D5F7D3C1107A0 CRC64;
CC
CC Query Match 5.3%; Score 93.5; DB 1; Length 1380;
CC Best Local Similarity 21.0%; Pred. No. 7.8;
CC Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;
CC
CC 3 GGRDAILTCAVHP---ELIFDITK-----LLAFGLPM-----V 36
CC DB 101 GGRDAILTCAVHP---ELIFDITK-----LLAFGLPM-----V 36
CC
CC QY 37 LQAGITKVP-----YFPAQGLIRACMLVRKXAGHYV-----QMAFVK 75
CC DB 156 VQSGMLILAPFTGSSGVRITMSDSVYFTRAEPVVELKVLN-----HIVRLRAREVAFVR 210
CC
CC QY 76 LAAITG-----YYVYDHLTPLODMAHAGRLAVA-VEVITSDMEYK----- 117
CC DB 211 -----LFGMHFGSEELTYVDTLTSL-----LRDPAVLTYVYSSSVVEDEAFDAMAD 260
CC
CC QY 118 -----IITWADTAACGDIISGLPVSARRGREILG-----PADNFEQ 156
CC DB 261 TNPQVITVAAPVQVITFLEKVLDPRTSSAYVISCMIQKIVPVDYKRLLSASISIRP 320
CC
CC QY 157 GKRLLAPITAYSOQTRGILGCIITSLNGRDKQVEG---VQVSTATQSLATCVNGV 212
CC DB 321 DGRILIASAT-----SPVSGEGIKVMEVILAKAMSNYIE---NSG 356
CC
CC QY 213 CWTVPFHAGSKTLAPK---GPITOMYTNVDQDLVGQAPGASMTPTCGSSDLYLV 268
CC DB 357 SFDDYPPDDSTLTGKRKARSEAPLSRKT-VDE---FQAHSTIKMALGLSGTLVQ 412
CC
CC QY 269 TRHADVTPYRRRGRSGSLSPRPV---SYKSSGGPILC-PSGHAVGIFRAVCTRGV 324
CC DB 413 TLEQTDWIVNRETRYKAGLFNQRRFVIGDYVLDYGGGP--CEPLAQPLG--ASCYCNQGG 468
CC
CC QY 325 AKAV-----DPIVPSME 337
CC DB 469 HSSITLVQNASMDLVDPDSFK 490
```

D5	290	GIDVSPFTFSPDGSMTAFVSTRGSP	314
RESULT 15			
TRFE HORSE			
ID	TRFE HORSE	STANDARD;	PRF; 706 AA.
AC	P27425;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin).		
GN	TF.		
OS	Equus caballus (Horse).		
OC	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus.		
OX	NCBI_TaxId=9796;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=933277958; PubMed=8504171;		
RT	Carpenter M.A., Broad T.E., transferrin";		
RL	"The cDNA sequence of horse transferrin";		
RL	Biochim. Biophys. Acta 1173:230-232(1993).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Extraembryonic tissue;		
RA	McPownell K.J., Adams M.H., Baker C.B.;		
RL	Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: Transferrins are iron binding transport proteins which can bind two atoms of ferric iron in association with the binding of an anion, usually bicarbonate. It is responsible for the transport of iron from sites of absorption and heme degradation to those of storage and utilization. Serum transferrin may also have a further role in stimulating cell proliferation.		
CC	-1- SUBUNIT: Monomer.		
CC	-1- SUBCELLULAR LOCATION: Secreted.		
CC	-1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.		
CC	-1- DOMAIN: Composed of two homologous domains.		
CC	-1- SIMILARITY: Belongs to the transferrin family.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL, M69020, AAA30958.1; -		
DR	EMBL, U21127, AAA63684.1; -		
DR	PIR, S33761; S33761.		
DR	HSSP, P02787; 1ABE.		
DR	InterPro, IPR001156; Transferrin.		
DR	Pfam, PF00405; transferrin; 2.		
DR	PRINTS, PR00422; TRANSFERRIN.		
DR	SMART, SM00094; TR_PIR_2		
DR	PROSITE, PS00205; TRANSFERRIN_1; 2.		
DR	PROSITE, PS00206; TRANSFERRIN_2; 2.		
DR	PROSITE, PS00207; TRANSFERRIN_3; 2.		
KW	Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;		
KW	Signal.		
FT	SIGNAL	1	19 BY SIMILARITY.
FT	CHAIN	20	706 SEROTRANSFERRIN.
FT	REPEAT	20	357 1.
FT	REPEAT	358	706 2.
FT	DISULFID	26	64 BY SIMILARITY.
FT	DISULFID	36	55 BY SIMILARITY.
FT	DISULFID	134	215 BY SIMILARITY.
FT	DISULFID	174	190 BY SIMILARITY.
FT	DISULFID	177	198 BY SIMILARITY.
FT	DISULFID	187	200 BY SIMILARITY.
FT	DISULFID	248	262 BY SIMILARITY.
FT	DISULFID		



```
FT DISULFID 360 623 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 376 389 BY SIMILARITY.
FT DISULFID 423 701 BY SIMILARITY.
FT DISULFID 441 664 BY SIMILARITY.
FT DISULFID 474 550 BY SIMILARITY.
FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MM; 1A0FA566C0409D8A CRC64;
```

Query Match 5.2%; Score 92.5; DB 1; Length 706;

Best Local Similarity 21.5%; Pred. No. 4.2; Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

```
QY 71 MAFKALATGTYY---YDHLTPLODWAHAGRLDVAVEVIFSDMEVKIITWGA----123
Db 321 LGFLRIPAMDWTWYLGVEYVT-----AIRNLREDIREVPKD-ECKVKWCAIGHH 371
QY 124 DTAACGD--ITSGLPVBARGR-----EILGPADNFEQGWRL-----LAPITAY 167
Db 372 EKVKCDSESVNSGNIECESAOSTEDCIATKIVKGEADAMSIDGGFIYIAGKGLVPLAE 431
QY 168 SOOTRGLGCIITSLTGSDKNQVEGEVQVSTATOSFIATCVNGVCTVPHGAGSKTIAG 227
Db 432 NYETRRSGSACVDTPEEGH-----AAVAVKSSSDPDLT-----W-----NSLKG 470
QY 228 PKGPTQMYTNVDQDLVGMQAPPARASMTPTCGSSDLYLVTRHADVIVPARRGDSRGL 287
Db 471 KK---SCHTGVDR-TAGWNIPLMGL-----LYSEIKHCEPDKFFREGCAPGYR 513
QY 288 LSPRPVSLKSSGGP-LIC-PSGHA-----VGFRAVCTRGVAKAVDFIPVSMSE 337
Db 514 RNSTLCNLCIGSASGPGRECEPNNHRYGYTGAFRLCLVKEGDVA---FVKHQTVE 566
```

Search completed: May 6, 2004, 09:31:54  
Job time : 8.119 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 / Search time 31.8931 Seconds

(without alignments)  
3373.509 Million cell updates/sec

Title: US-10-650-585-14  
Perfect score: 1778  
Sequence: 1 VFGGRDAIILLTCAVPELI.....RGVAKAVDFIPVESMETWR 341

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPTRMBL\_25:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1729	97.2	3010	12 Q9DTB6	Q9dtb6 hepatitis c
2	1726	97.1	3010	12 Q9J3F9	Q9j3f9 hepatitis c
3	1721	96.8	3010	12 Q9DTD6	Q9dtd6 hepatitis c
4	1719	96.7	3010	12 Q9J3H7	Q9j3h7 hepatitis c
5	1719	96.7	3010	12 P90191	P90191 hepatitis c
6	1718	96.6	3010	12 Q9J3H5	Q9j3h5 hepatitis c
7	1715	96.5	3008	12 Q9J3F4	Q9j3f4 hepatitis c
8	1715	96.5	3010	12 Q9J3H3	Q9j3h3 hepatitis c
9	1715	96.5	3010	12 Q9DTD7	Q9dtd7 hepatitis c
10	1715	96.5	3010	12 Q807P3	Q807p3 hepatitis c
11	1714	96.4	3010	12 P88803	P88803 hepatitis c
12	1711	96.2	1186	12 Q81755	Q81755 hepatitis c
13	1711	96.2	2284	12 Q81817	Q81817 hepatitis c
14	1711	96.2	3010	12 Q68788	Q68788 hepatitis c
15	1711	96.2	3010	12 P89366	P89366 hepatitis c
16	1710	96.2	3010	12 Q99A02	Q99a02 hepatitis c

17	1710	96.2	3010	12 Q9QIX6	Q9qix6 hepatitis c
18	1710	96.2	3010	12 Q9QIX5	Q9qix5 hepatitis c
19	1710	96.2	3010	12 Q68826	Q68826 hepatitis c
20	1710	96.2	3010	12 Q9J3G6	Q9j3g6 hepatitis c
21	1710	96.2	3010	12 Q9DTB4	Q9dtb4 hepatitis c
22	1709	96.1	3010	12 Q9J3H9	Q9j3h9 hepatitis c
23	1708	96.1	3010	12 Q9J3I0	Q9j3i0 hepatitis c
24	1708	96.1	3010	12 Q09796	Q09796 hepatitis c
25	1708	96.1	3011	12 Q9DTB3	Q9dtb3 hepatitis c
26	1707	96.0	3010	12 Q9QIY3	Q9qi y3 hepatitis c
27	1705	95.9	3010	12 Q9QIY5	Q9qi y5 hepatitis c
28	1705	95.9	3014	12 Q9DTB0	Q9dtb0 hepatitis c
29	1704	95.8	3010	12 Q9DTF0	Q9dtf0 hepatitis c
30	1703	95.8	3010	12 Q02829	Q02829 hepatitis c
31	1702	95.7	3010	12 Q9J3H0	Q9j3h0 hepatitis c
32	1702	95.7	3010	12 Q81541	Q81541 hepatitis c
33	1702	95.7	3010	12 Q9WMX2	Q9wm x2 hepatitis c
34	1701	95.7	3010	12 Q81760	Q81760 hepatitis c
35	1701	95.7	3010	12 Q9QIX8	Q9qix8 hepatitis c
36	1701	95.7	3010	12 Q9QIX7	Q9qix7 hepatitis c
37	1701	95.7	3010	12 Q8QRL8	Q8qrl8 hepatitis c
38	1701	95.7	3010	12 Q9J3H6	Q9j3h6 hepatitis c
39	1699	95.6	3010	12 Q9J3I1	Q9j3i1 hepatitis c
40	1698	95.5	3010	12 Q9J3H2	Q9j3h2 hepatitis c
41	1698	95.5	3010	12 Q9DTB5	Q9dtb5 hepatitis c
42	1698	95.5	3010	12 Q9DTB7	Q9dtb7 hepatitis c
43	1698	95.5	3010	12 Q9QIY4	Q9qi y4 hepatitis c
44	1697	95.4	3010	12 Q9IAU0	Q9ia u0 hepatitis c
45	1697	95.4	3010	12 Q9J3G3	Q9j3g3 hepatitis c

## ALIGNMENTS

RESULT 1  
Q9DTB6 PRELIMINARY; PRT; 3010 AA.  
AC Q9DTB6;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HCV142;  
RA Takahashi K., Iwata K., Matsumoto H., Nakao K.,  
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hikioka M.,  
RA Mishiro S.;  
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
RT with hepatocellular carcinoma: the 'progression score' revisited.";  
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA (BY SIMILARITY).  
CC EMBL; AB049091; BAB1804.1; -;  
DR PIR; A61196; A61196.  
DR PIR; PS0329; PS0329.  
DR HSSP; P28663; LUXP.  
DR GO; GO:0016021; C: integral to membrane; IEA.  
DR GO; GO:0019828; C: viral capsid; IEA.  
DR GO; GO:0019031; C: viral envelope; IEA.  
DR GO; GO:0005524; F: ATP binding; IEA.  
DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.  
DR GO; GO:0005489; F: electron transporter activity; IEA.  
DR GO; GO:0016787; F: hydrolase activity; IEA.  
DR GO; GO:0003723; F: RNA binding; IEA.  
DR GO; GO:0003668; F: RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F: serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral genome replication; IEA.  
 DR InterPro: IPR009003; Cys Ser lysozyme.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C2.  
 DR InterPro: IPR007095; RNA pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD18662; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC\_C; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolyase; Nonstructural protein; Polypeptide;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 97.2%; Score 1729; DB 12; Length 3010;  
 Best Local Similarity 95.9%; Pred. No. 1,36-138;  
 Matches 327; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

DR 1 VRGRDAIILITCAVHELIIFDITKILAIIFGPIIMVLAQGITVVFVRAQGLIRACMLV 60  
 866 VRGRDAIILITCAVHELIIFDITKILAIIFGPIIMVLAQGITVVFVRAQGLIRACMLV 925

DR 61 RKAAGHYVQMAFKALALITGYVDHLTLPQMAHAGLDAVAVVVFSEMEYKILIT 120  
 926 RKAAGHYVQMAFKALALITGYVDHLTLPQMAHAGLDAVAVVVFSEMEYKILIT 985

DR 121 WGAADTAACGDIISGLPVSARSGREIILGPADNFGSGMRILAPITAVSOOTRILACIIT 180  
 986 WGAADTAACGDIISGLPVSARSGREIILGPADNFGSGMRILAPITAVSOOTRILACIIT 1045

DR 181 SLTRDXNQVGEVQVSTATQSFPLATCVGCVTFVHAGSKTLGPKPITQMTYND 240  
 1046 SLTRDXNQVGEVQVSTATQSFPLATCVGCVTFVHAGSKTLGPKPITQMTYND 1105

DR 241 QDLVGMQAPPGARSMCTCGSSDLVLTTHADVTPRRSGDRGSLISRPVSYSKGS 300  
 1106 QDLVGMQAPPGARSMCTCGSSDLVLTTHADVTPRRSGDRGSLISRPVSYSKGS 1165

DR 301 GGPLCPSGHAYGIFRAVCTRGAKAVDPIVSEMETTNR 341  
 1166 GGPLCPSGHAYGIFRAVCTRGAKAVDPIVSEMETTNR 1206

RESULT 2

09J3F9 ID 09J3F9 PRELIMINARY; PRT; 3010 AA.  
 AC 09J3F9  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NC NCB1\_Taxid=11103;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD3;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease progression."  
 RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: AF207774; AAF65964.1; -  
 DR PIR: A61196; A61196.  
 DR PIR: P00246; P00246.  
 DR PIR: P03329; P03329.  
 DR HSSP: P27958; 1HE1.  
 DR MEROPS: S29.001; -  
 DR GO:0016021; C:integral to membrane; IEA.  
 DR GO:0019028; C:viral capsid; IEA.  
 DR GO:0018031; C:viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008235; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; P:electron transport; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral genome replication; IEA.  
 DR InterPro: IPR009003; Cys Ser lysozyme.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C2.  
 DR InterPro: IPR007095; RNA pol\_D5\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD18662; HCV\_NS1; 1.

DR SMART, SM00487, DEXDC, 1.  
 DR PROSITE; PS00190; CYTOCHROME C, 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Glycase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C0DB93E60C7 CRC64;

Query Match 97.1%; Score 1726; DB 12; Length 3010;  
 Best Local Similarity 95.6%; Pred. No. 2,4e-138;  
 Matches 326; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 1 VRGGRDAIILTCVHPELIFDITKLLAFGLPMVLAQITKPYFVRAQGLIRACMLV 60  
 DB 866 VRGGRDAIILTCVHPELIFDITKLLAFGLPMVLAQITKPYFVRAQGLIRACMLV 925  
 QY 61 RKAAGHYVQMAFMKLAALGTYYDHLTPLOMAHAGRLDAVAEPIFSMEVKIIT 120  
 DB 926 RKAAGHYVQMAFMKLAALGTYYDHLTPLOMAHAGRLDAVAEPIFSMEVKIIT 985  
 QY 121 WGAUTAACGDIISGLPVSARGRRIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 180  
 DB 986 WGAUTAACGDIISGLPVSARGRRIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 1045  
 QY 181 SLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKGPITQMTYND 240  
 DB 1046 SLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKGPITQMTYND 1105  
 QY 241 QDLVGMQAPPGARSMPTCTGSSDLVYTRHADVIPIRRRGRSGSLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMPTCTGSSDLVYTRHADVIPIRRRGRSGSLSPRPVSYLKSS 1165  
 QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 341  
 DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 3  
 Q9DND6 PRELIMINARY; PRT; 3010 AA.  
 ID Q9DND6  
 AC Q9DND6  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OS Hepatitis C virus.  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 CX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HCV221;  
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Hatahara T., Ohra Y., Kanai K., Maruo H., Baba K., Hijioka M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients  
 with hepatocellular carcinoma: the progression score revisited.";  
 RL Submitted (Sep-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MENA (BY SIMILARITY).  
 CC EMBL; AB049101; BAB18814.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; PQ0246; PQ0246.  
 DR PIR; PS0329; PS0329.  
 DR HSSP; P2663; 1UXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0016787; F:helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_typsin.  
 DR InterPro; IPR001345; Cys\_heme\_BS.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002511; HCV\_NSI.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PsVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRp; 1.  
 DR ProDom; ED186062; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; PS00190; CYTOCHROME C, 1.  
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydroxylase; Nonstructural protein; Polyprotein;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327108 MW; DE182D810E78E84 CRC64;

Query Match 96.8%; Score 1721; DB 12; Length 3010;  
 Best Local Similarity 95.9%; Pred. No. 6.5e-138;  
 Matches 327; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGRDAIILTCVHPELIFDITKLLAFGLPMVLAQITKPYFVRAQGLIRACMLV 60  
 DB 866 VRGGRDAIILTCVHPELIFDITKLLAFGLPMVLAQITKPYFVRAQGLIRACMLV 925  
 QY 61 RKAAGHYVQMAFMKLAALGTYYDHLTPLOMAHAGRLDAVAEPIFSMEVKIIT 120  
 DB 926 RKAAGHYVQMAFMKLAALGTYYDHLTPLOMAHAGRLDAVAEPIFSMEVKIIT 985  
 QY 121 WGAUTAACGDIISGLPVSARGRRIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 180  
 DB 986 WGAUTAACGDIISGLPVSARGRRIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 1045  
 QY 181 SLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKGPITQMTYND 240  
 DB 1046 SLTGRDNQVGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKGPITQMTYND 1105  
 QY 241 QDLVGMQAPPGARSMPTCTGSSDLVYTRHADVIPIRRRGRSGSLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMPTCTGSSDLVYTRHADVIPIRRRGRSGSLSPRPVSYLKSS 1165  
 QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 341  
 DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 4

Q9J3H7 PRELIMINARY; PRT; 3010 AA.

AC Q9J3H7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI\_Taxid=11103;

RE SEQUENCE FROM N.A.

RC STRAIN=MD5; Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.,

RA Miyasaka K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.,

RT "Characteristics of hepatitis C viral genome associated with disease progression."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

EMBL; AF20756; AAF65946.1; -

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00804; P00804.

DR PIR; PS0329; PS0329.

DR HSSP; P26663; IUXP.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008266; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006118; F:electron transport; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P:transcription; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR GO; GO:0019087; P:viral transformation; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR00345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NSI.

DR InterPro; IPR002518; HCV\_NSI.

DR InterPro; IPR000745; HCV\_NS4A.

DR InterPro; IPR001490; HCV\_NS4B.

DR InterPro; IPR002868; HCV\_NS5A.

DR InterPro; IPR002168; HCV\_NS5B.

DR InterPro; IPR004109; peptidase\_C29.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NSI; 1.

DR Pfam; PF01538; HCV\_NSI; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4; 1.

DR Pfam; PF01001; HCV\_NS4B; 1.

DR Pfam; PF01506; HCV\_NS5A; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR ProDom; PD186062; HCV\_NSI; 1.

DR SMART; SM00487; DEXDC; 1.

DR PROSITE; PS00190; CYTOCHROME\_C; 1.

KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polyprotein; RNA-directed RNA polymerase; transferase; Transmembrane.

SC SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Query Match 96.7%; Score 1719; DB 12; Length 3010;

Best Local Similarity 95.0%; Pred. No. 9.6e-138;

Matches 324; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 VGGSDAIIILTCVAHPELLFDITLTLAIGPLMTLOAGITKVPYFPAAGLIRACMLV 60

DB 866 VGGSDAIIILTCVAHPELLFDITLTLAIGPLMTLOAGITKVPYFPAAGLIRACMLV 925

QY 61 RKAGGHVQVQAFMLTALITGVYVDHLTPLOVMAHAGLDAVAVEPIVSDMEVKIT 120

DB 926 RKAGGHVQVQAFMLTALITGVYVDHLTPLOVMAHAGLDAVAVEPIVSDMEVKIT 985

QY 121 WQADTAACGDIISGLPVASRRREILGPDNPFEGQGRLLAPITAYSQOTRGLGCIIT 180

DB 986 WQADTAACGDIISGLPVASRRREILGPDNPFEGQGRLLAPITAYSQOTRGLGCIIT 1045

QY 181 SLTGRDNQVGEVQVSTATOSPLATGVNVCMTVPFGAGSKTLAGKPIITOMYTNVD 240

DB 1046 SLTGRDNQVGEVQVSTATOSPLATGVNVCMTVPFGAGSKTLAGKPIITOMYTNVD 1105

QY 241 QDLVGMQAPFGARSMTPCTCGSSDLYLVTRHADYIPVRRDSSGSLSPRVSYLKSS 300

DB 1106 QDLVGMQAPFGARSMTPCTCGSSDLYLVTRHADYIPVRRDSSGSLSPRVSYLKSS 1165

QY 301 GGPLICPSGHAVGIFRAVCTRGVAKAVDFIPVSMETMTM 341

DB 1166 GGPLICPSGHAVGIFRAVCTRGVAKAVDFIPVSMETMTM 1206

RESULT 5

P90191 PRELIMINARY; PRT; 3010 AA.

AC P90191;

DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI\_Taxid=11103;

RE SEQUENCE FROM N.A.

RC STRAIN=HCV-1b;

RA Enomoto N.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=HCV-1b;

RX MEDLINE=95340824; PubMed=7542279;

RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,

RA Yamamoto C., Izumi N., Marumo F., Sato C.,

RT "Comparison of full-length sequences of interferon-sensitive and

RT resistant hepatitis C virus 1b."

RT J. Clin. Invest. 96:224-230(1995).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

EMBL; D50482; BAA09073.1; -

DR PIR; A61196; A61196.

DR PIR; P00254; P00254.

DR PIR; P00804; P00804.

DR PIR; PS0329; PS0329.

DR PDB; 1DXW; 12-JAN-01.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0001674; F:transferase activity; IEA.  
 DR GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_typsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NSI.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR001490; HCV\_NS4a.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NSI; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR Pfam: PD186062; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR Coar protein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 KW Polypeptide; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 FT CHAIN 1 191 CORE PROTEIN.  
 FT CHAIN 192 383 E1.  
 FT CHAIN 384 809 E2.  
 FT CHAIN 810 1026 NS2.  
 FT CHAIN 1027 1657 NS3.  
 FT CHAIN 1658 1711 NS4a.  
 FT CHAIN 1712 1972 NS4b.  
 FT CHAIN 1973 2419 NS5a.  
 FT CHAIN 2420 3010 NS5b.  
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C268 CRC64;

Query Match 96.7%; Score 1719; DB 12; Length 3010;  
 Best Local Similarity 95.3%; Pred. No. 9,66-138;  
 Matches 325; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 1 VRGGRDAILLTCAVHPELITDITKLLAIFGLMTLOAGITKPYVRAAGLIRACMLV 60  
 DB 866 VRGGRDAILLTCAVHPELITDITKLLAIFGLMTLOAGITKPYVRAAGLIRACMLV 925  
 QY 61 RKAAGHVVQMAFMKLAALGTYYVVDHLPDQMAHAGLRLAVAVEPFIQSDMEVKIT 120  
 DB 926 RKAAGHVVQMAFMKLAALGTYYVVDHLPDQMAHAGLRLAVAVEPFIQSDMEVKIT 985  
 QY 121 WGADTAACGDIISGLIPVARGREILIGPADNFEQGGRLIAPITAVSQOTRGLIGCII 180  
 DB 986 WGADTAACGDIISGLIPVARGREILIGPADNFEQGGRLIAPITAVSQOTRGLIGCII 1045  
 QY 181 SLTGRDKNQGVGEVQVSTATOSFLATGVNGVCMVTFVAGASKTLGKGIOTMYNVD 240  
 DB 1046 SLTGRDKNQGVGEVQVSTATOSFLATGVNGVCMVTFVAGASKTLGKGIOTMYNVD 1105  
 QY 241 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVI PVRRRGDSRGSLLSPRPVSYLKGS 300

DB 1106 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVI PVRRRGDSRGSLLSPRPVSYLKGS 1165  
 QY 301 GGPRLCPSGHANGIFRANCTGVAKAVDFIVESMETTMR 341  
 DB 1166 GGPRLCPSGHANGIFRANCTGVAKAVDFIVESMETTMR 1206  
 RESULT 6  
 ID Q9J3H5 PRELIMINARY; PRT; 3010 AA.  
 AC Q9J3H5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD17;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease progression."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL; AF207758; AAF65948.1; -  
 DR PIR; A61196; A61196.  
 DR PIR; P00246; P00246.  
 DR PIR; P00254; P00254.  
 DR PIR; P50329; P50329.  
 DR HSSP; P27938; IHE1.  
 DR GO:0016021; C:Integral to membrane; IEA.  
 DR GO:0019028; C:Viral capsid; IEA.  
 DR GO:0019031; C:Viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0005489; F:electron transporter activity; IEA.  
 DR GO:0016787; F:hydrolase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006118; F:electron transport; IEA.  
 DR GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_typsin.  
 DR InterPro: IPR000345; Cys\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NSI.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS5a.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NSI; 1.

QY	DB	Query Match	Best Local Similarity	Matches	326; Conservative	Score 1715; DB 12; Length 3008;	Pred. No. 2.1e-117;	Mismatches 9;	Indels 0;	Gaps 0
QY	1	VRRGRPAITLLCAVAHPELIPITLLAIFGPMVLQAGITKPYFVAQGLIRACMLV	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
DB	864	VRRGRPAITLLCAVAHPELIPITLLAIFGPMVLQAGITKPYFVAQGLIRACMLA	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
QY	61	RKAAGHYVQMAFMKIALTGTVYVDHLPLQDMAHGLRDLAVAEVIFSMENVKIIT	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
DB	924	RKAAGHYVQMAFMKIALTGTVYVDHLPLQDMAHGLRDLAVAEVIFSMENVKIIT	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
QY	121	WGADTLACGDIITSGIPVABRRREITLGPANFEGGOMRLAPITAYSQOTRGLGCIIT	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
DB	984	WGADTLACGDIITSGIPVABRRREITLGPANFEGGOMRLAPITAYSQOTRGLGCIIT	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
QY	181	SLTGRDKNVVEEYVNVSTATQSFATCVGVCMTVFHGAQSKTLAIGPKSPITQMTYND	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
DB	1044	SLTGRDKNVVEEYVNVSTATQSFATCVGVCMTVFHGAQSKTLAIGPKSPITQMTYND	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
QY	241	QDLVKGQAPPGKARMSPTCTCGSSDLIVLTRADVIPIRRGDSRGSLSPISYIKGSS	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0
DB	300	QDLVKGQAPPGKARMSPTCTCGSSDLIVLTRADVIPIRRGDSRGSLSPISYIKGSS	96.5%;	326;	6;	1715;	2.1e-117;	9;	0;	0



```

Db      1104 QDLVGMQAPPGARSLFPCTCGSSDLYLTVRHADVIPRRRGRSGSLSPRVSYLKSS 1163
QY      301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIVPESMETTMR 341
Db      1164 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIVPESMETTMR 1204

RESULT 8
ID      09JTH3 PRELIMINARY; PRT; 3010 AA.
AC      09JTH3;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MD19;
RA      Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT      "Characteristics of hepatitis C viral genome associated with disease
RT      progression.";
RL      Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA (BY SIMILARITY).
DR      EMBL; AF207760; AAF65950.1; -.
DR      PIR; A61196; A61196.
DR      HSSP; P26663; 1UXP.
DR      GO; GO:0016021; C:integral to membrane; IEA.
DR      GO; GO:0019028; C:viral capsid; IEA.
DR      GO; GO:0019031; C:viral envelope; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR      GO; GO:0005489; F:electron transporter activity; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR      GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      GO; GO:0006350; P:transcription; IEA.
DR      GO; GO:0019079; P:viral genome replication; IEA.
DR      GO; GO:0019087; P:viral transformation; IEA.
DR      InterPro; IPR008003; Cys_Ser_cryptin.
DR      InterPro; IPR000345; CytC_heme_35.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR02522; HCV_capsid.
DR      InterPro; IPR02521; HCV_core.
DR      InterPro; IPR002519; HCV env.
DR      InterPro; IPR002531; HCV_NS1.
DR      InterPro; IPR002518; HCV_NS2.
DR      InterPro; IPR000745; HCV_NS4a.
DR      InterPro; IPR001490; HCV_NS4b.
DR      InterPro; IPR002868; HCV_NS5a.
DR      InterPro; IPR002166; HCV_RdRp.
DR      InterPro; IPR001650; Helicase_C.
DR      InterPro; IPR004109; peptidase_C9.
DR      InterPro; IPR007095; RNA_pol_DS_PS.
DR      InterPro; IPR007094; RNA_pol_PSVLr.
DR      Pfam; PF01543; HCV_capsid; 1.
DR      Pfam; PF01542; HCV_core; 1.
DR      Pfam; PF01539; HCV_env; 1.
DR      Pfam; PF01560; HCV_NS1; 1.
DR      Pfam; PF01538; HCV_NS2; 1.
DR      Pfam; PF02907; HCV_NS3; 1.

```

```

DR      Pfam; PF01006; HCV_NS4a; 1.
DR      Pfam; PF01001; HCV_NS4b; 1.
DR      Pfam; PF01506; HCV_NS5a; 1.
DR      Pfam; PF00271; helicase_C; 1.
DR      Pfam; PF00998; viral_RdRp; 1.
DR      ProDom; PD186062; HCV_NS1; 1.
DR      SMART; SM00487; DEXdc; 1.
DR      PROSITE; PS00190; CYTOCHROME_C_1.
DR      Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;
DR      Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
DR      POLYPROTEIN; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3010 AA; 327234 MW; 44C34677649CB8DD CRC64;

Query Match      96.5%; Score 1715; DB 12; Length 3010;
Best Local Similarity 94.4%; Pred. No. 2,1e-137;
Matches 322; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

QY      1 VRGGRDAITLTCVAHPDLIPITKLLAIFGLMVLQAGITKVPYFRAQGLIRACMLV 60
Db      866 VAGGRDAITLTCVAHPDLIPITKLLAIFGLMVLQAGITRVPYFRAQGLIRACMLA 925
QY      61 RRAAGHYVQMAFKLALITGYVDHITPLQDAAHAGLRDLAVAEVITSDEVKIIT 120
Db      926 RRVAGGHYVQMAFKLALITGYVDHITPLQDMAHAGLRDLAVAEVITSDMETKIIT 985
QY      121 WQADPAAAGDIISGLPVASRRGRITLGPADNFEQGRLLAPITAYSQOTRGLGCIIT 180
Db      986 WQADPAAAGDIISGLPVASRRGRITLGPADNFEQGRLLAPITAYSQOTRGLGCIIT 1045
QY      181 SLTGRDNQVEGEVQVSTATQSFATCVNGVCTVFHAGSKTLAGEKPIITQYITVD 240
Db      1046 SLTGRDNQVEGEVQVSTATQSFATCVNGVCTVFHAGAKTLAGEKPIITQYITVD 1105
QY      241 QDLVGMQAPPGARSLFPCTCGSSDLYLTVRHADVIPRRRGRSGSLSPRVSYLKSS 300
Db      1106 QDLVGMQAPPGARSLFPCTCGSSDLYLTVRHADVIPRRRGRSGSLSPRVSYLKSS 1165
QY      301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIVPESMETTMR 341
Db      1166 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIVPESMETTMR 1206

RESULT 9
ID      09DTH7 PRELIMINARY; PRT; 3010 AA.
AC      09DTH7;
DT      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=HCV1217;
RA      Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA      Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA      Mishiro S.;
RT      "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT      with hepatocellular carcinoma: the 'progression score' revisited.";
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA (BY SIMILARITY).
DR      EMBL; AB049100; BAB18813.1; -.
DR      PIR; A61196; A61196.
DR      PIR; PQ0804; PQ0804.
DR      PIR; PS0329; PS0329.
DR      HSSP; P26663; 1UXP.
DR      MEROPS; S29.002; -.
DR      GO; GO:0016021; C:integral to membrane; IEA.

```

DR GO: 00019028; C: viral capsid; IEA.  
 DR GO: 00019031; C: viral envelope; IEA.  
 DR GO: 00005524; F: ATP binding; IEA.  
 DR GO: 00008026; F: ATP dependent helicase activity; IEA.  
 DR GO: 00005489; F: electron transporter activity; IEA.  
 DR GO: 00003723; F: RNA binding; IEA.  
 DR GO: 00003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: 00008236; F: serine-type peptidase activity; IEA.  
 DR GO: 00005188; F: structural molecule activity; IEA.  
 DR GO: 00016740; F: transferase activity; IEA.  
 DR GO: 00006118; F: electron transport; IEA.  
 DR GO: 00006508; F: proteolysis and peptidolysis; IEA.  
 DR GO: 00006350; F: transcription; IEA.  
 DR GO: 00019079; P: viral genome replication; IEA.  
 DR GO: 00019087; P: viral transformation; IEA.  
 DR InterPro: IPR003003; Cys Ser trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR004109; Peptidase C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase C; 1.  
 DR Pfam: PF00998; viral RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326909 MW; 5505C62EB0DA0519 CRC64;

Query Match 96.58; Score 1715; DB 12; Length 3010;  
 Best Local Similarity 95.38; Pred. No. 2,1e-137;  
 Matches 325; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

DR 1 VRGGRDAIILITCAVHPELFDITKLLAFGLPLVLAAGITKVPFVBAQGLIRACMLV 60  
 DB 866 VRGGRDAIILITCAVHPELFDITKLLAFGLPLVLAAGITKVPFVBAQGLIRACMLV 925  
 QY 61 RKAAGHYVQAFMKLAALTGTYYVDHLTPLOMANAGLDLAANVAVPVSMEVXIIT 120  
 DB 926 RKAAGHYVQAFMKLAALTGTYYVDHLTPLOMANAGLDLAANVAVPVSMEVXIIT 985  
 QY 121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGOQWRLAITYSQOTRGLLCIT 180  
 DB 986 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGOQWRLAITYSQOTRGLLCIT 1045  
 QY 181 SLTGRDNQVEGEVQVSTATOSFLATCVNGVQVTFHAGSKTLAGPKPIITOMYTNV 240  
 DB 1046 SLTGRDNQVEGEVQVSTATOSFLATCVNGVQVTFHAGSKTLAGPKPIITOMYTNV 1105  
 QY 241 QDLVNGQAPRGARSMCTCGSSDLYIVPHADIVVRRRGDSRGLSPRPVYKGS 300  
 DB 1106 QDLVNGQAPRGARSMCTCGSSDLYIVPHADIVVRRRGDSRGLSPRPVYKGS 1165

QY 301 GSELLCPSGHVGIFPRAVCTRGVAKAVDFIPVESMETMR 341  
 DB 1166 GSELLCPSGHVGIFPRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 10  
 ID 080723 PRELIMINARY; PRT; 3010 AA.  
 AC 080723;  
 DT 01-JUN-2003 (Trembl) 24, Created)  
 DT 01-JUN-2003 (Trembl) 24, Last sequence update)  
 DT 01-OCT-2003 (Trembl) 25, Last annotation update)  
 DE Polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OK NCBI:taxid=11103;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MLE;  
 RX MEDLINE=22047193; PubMed=12051758;  
 RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,  
 RA Nio Y., Hosaka K., Miyazaki Y., Shimotohno K.,  
 RT "Subgenomic replicon derived from a cell line infected with the  
 RT hepatitis C virus.",  
 RT Hepatitis C virus.",  
 RT Biochem. Biophys. Res. Commun. 293:993-999(2002).  
 RL EMBL, AB080299; BAC54896.1; -;  
 DR GO: 00019028; C: viral capsid; IEA.  
 DR GO: 00019031; C: viral envelope; IEA.  
 DR GO: 00005524; F: ATP binding; IEA.  
 DR GO: 00008026; F: ATP dependent helicase activity; IEA.  
 DR GO: 00005489; F: electron transporter activity; IEA.  
 DR GO: 00003723; F: RNA binding; IEA.  
 DR GO: 00003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: 00008236; F: serine-type peptidase activity; IEA.  
 DR GO: 00006350; F: transcription; IEA.  
 DR GO: 00006118; F: proteolysis and peptidolysis; IEA.  
 DR GO: 00005188; F: structural molecule activity; IEA.  
 DR GO: 00006508; F: electron transport; IEA.  
 DR GO: 00006118; F: proteolysis and peptidolysis; IEA.  
 DR GO: 00006350; F: transcription; IEA.  
 DR GO: 00019079; P: viral genome replication; IEA.  
 DR GO: 00019087; P: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR004109; Peptidase C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase C; 1.  
 DR Pfam: PF00998; viral RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; 1.  
 DR Polyprotein.

SEQUENCE 3010 AA; 327097 MW; EE6418C7A23E66 CRC64;  
 Query Match 96.5%; Score 1715; DB 12; Length 3010;  
 Best Local Similarity 95.6%; Pred. No. 2.1e-137;  
 Matches 326; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGRDAILLTCAVHPELIDITKLLAIFGPELWVLOAGTKVYFVRAOGLIRACMLV 60  
 DB 866 IRGRDAILLTCAVHPELIDITKLLAIFGPELWVLOAGTKVYFVRAOGLIRACMLV 925  
 QY 61 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVFSDMEVKIT 120  
 DB 926 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVFSDMEVKIT 985  
 QY 121 WGAADTAACGDIISGLPVASRRREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 180  
 DB 986 WGAADTAACGDIISGLPVASRRREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 1045  
 QY 181 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCFVHAGSKTLAPKGPITQMYTNVD 240  
 DB 1046 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCFVHAGSKTLAPKGPITQMYTNVD 1105  
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADYIPVRRRDSRGLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADYIPVRRRDSRGLSPRPVSYLKSS 1165  
 QY 301 GGPLICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 DB 1166 GGPLICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 11  
 P88803 PRELIMINARY; PRT: 3010 AA.  
 ID P88803  
 AC P88803  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DS Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=HCV-1b;  
 RX MEDLINE=95340824; PubMed=7542279;  
 RA Yamamoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,  
 Yamamoto C., Izumi N., Marumo F., Sato C.;  
 RT "Comparison of full-length sequences of interferon-sensitive and  
 RT resistant hepatitis C virus 1b."  
 RU J. Clin. Invest. 96:224-230 (1995).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA (BY SIMILARITY).  
 DR EMBL: D50484; BAA09075.1; -.  
 DR PIR: A61196; A61196.  
 DR HSP: P26663; INS3.  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.

DR GO: GO:0016740; F: transferase activity; IEA.  
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; F: transcription; IEA.  
 DR GO: GO:0019079; F: viral genome replication; IEA.  
 DR GO: GO:0019087; F: viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_C29.  
 DR InterPro: IPR007035; RNA\_pol\_D5\_PS.  
 DR InterPro: IPR007034; RNA\_pol\_Pevir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02987; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR Coac protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 FT CHAIN 1 191  
 FT CHAIN 192 383  
 FT CHAIN 384 809  
 FT CHAIN 810 1026  
 FT CHAIN 1027 1657  
 FT CHAIN 1658 1711  
 FT CHAIN 1712 1972  
 FT CHAIN 1973 2419  
 FT CHAIN 2420 3010  
 FT CHAIN NS5B.  
 SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEF7B8 CRC64;

Query Match 96.4%; Score 1714; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 2.6e-137;  
 Matches 323; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 VRGRDAILLTCAVHPELIDITKLLAIFGPELWVLOAGTKVYFVRAOGLIRACMLV 60  
 DB 866 VAGSGDAIILTCVAHSEIVEITKLLAIFGPELWVLOAGTKVYFVRAOGLIRACMLV 925  
 QY 61 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVFSDMEVKIT 120  
 DB 926 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVFSDMEVKIT 985  
 QY 121 WGAADTAACGDIISGLPVASRRREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 180  
 DB 986 WGAADTAACGDIISGLPVASRRREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 1045  
 QY 181 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCFVHAGSKTLAPKGPITQMYTNVD 240  
 DB 1046 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCFVHAGSKTLAPKGPITQMYTNVD 1105  
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADYIPVRRRDSRGLSPRPVSYLKSS 300  
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADYIPVRRRDSRGLSPRPVSYLKSS 1165  
 QY 301 GGPLICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 DB 1166 GGPLICPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 1206



RA	MEDLINE=94068484; PubMed=7502483;
RA	Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,
RA	Akagi T., Kimura K., Shimotohno K.;
RT	"Proteolytic processing and membrane association of putative
RT	nonstructural proteins of hepatitis C virus.";
RT	[Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
RL	
RN	
RP	SEQUENCE FROM N.A.
RP	MEDLINE=9433810; PubMed=8056334;
RP	Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
RT	"Identification of the domain required for trans-cleavage activity of
RT	hepatitis C viral serine proteinase.";
RT	Gene 145:215-219(1994).
RL	
RN	
RP	SEQUENCE FROM N.A.
RP	MEDLINE=95156583; PubMed=7853491;
RP	Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.;
RT	"Hepatitis C virus-encoded nonstructural protein NS4A has versatile
RT	functions in viral protein processing.";
RT	J. Virol. 69:1575-1581(1995).
RL	
DR	EMB1; D16435; BAA03905.1; -.
DR	PIR; A61196; A61196.
DR	PIR; P00246; P00246.
DR	PIR; P80329; P80329.
DR	HSSP; P26663; IUXP.
DR	GO; GO:0019012; C:viralon; IEA.
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR	GO; GO:0003723; F:RNA binding; IEA.
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR	GO; GO:0016740; F:serine-type peptidase activity; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	GO; GO:0006350; P:transcription; IEA.
DR	GO; GO:0019079; P:viral genome replication; IEA.
DR	GO; GO:0019087; P:viral transformation; IEA.
DR	InterPro; IPR009003; Cys_ser_1trypsin.
DR	InterPro; IPR001410; DEAD.
DR	InterPro; IPR002518; HCV NS2.
DR	InterPro; IPR007455; HCV NS4A.
DR	InterPro; IPR001490; HCV NS4b.
DR	InterPro; IPR002868; HCV NS5a.
DR	InterPro; IPR002166; HCV_RdRp.
DR	InterPro; IPR001650; Helicase_C.
DR	InterPro; IPR004109; Peptidase_C9.
DR	InterPro; IPR007095; RNA_pol_DS_Ps.
DR	InterPro; IPR007094; RNA_pol_PSVir.
DR	Pfam; PF01538; HCV NS2; 1.
DR	Pfam; PF02907; HCV NS3; 1.
DR	Pfam; PF01006; HCV_NS4a; 1.
DR	Pfam; PF01001; HCV_NS4b; 1.
DR	Pfam; PF01506; HCV_NS5a; 1.
DR	Pfam; PF00271; helicase_C; 1.
DR	Pfam; PF00998; Viral_RdRp; 1.
DR	SMART; SMO0487; DEXDC; 1.
KW	Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;
KW	Signal; Transferrase.
FT	SIGNAL
FT	4
FT	55
FT	83
FT	21
FT	54
FT	84
FT	300
FT	931
FT	932
FT	985
FT	986
FT	1246
FT	1247
FT	1693
FT	NS5A.
FT	NS5B.
FT	1694
FT	2284

SEQ	SEQUENCE	2284 AA	247213 NM	DC272A1517046337 CRR64
Query Match		96.2%	Score 1711	DB 12; Length 2284;
Best Local Similarity		95.3%	Pred. No. 3.2e-137	
Matches	325; Conservative	9;	Mismatches 7;	Indels 0; Gaps 0;
QY	1 VRGGDAIILLTCVAHPFLIDITFLALAIQPLMLVLAQITTKYPIFYRAGGLIRACMLV	60		
DB	140 VRGRDAIILLTCVAHPFLIDITFLALAIQPLMLVLAQITTKYPIFYRAGGLIRACMLV	199		
QY	61 RKAAGHYVQVAFMKLAALITGYVYDHLTPQDMAHAGLRDLAAVEPVISDEVKIIT	120		
DB	200 RKAAGHYVQVAFMKLAALITGYVYDHLTPQDMAHAGLRDLAAVEPVISDEVKIIT	259		
QY	121 WGAIDTAACGDIISGLPVSAARGREIILGPANFEGCGRRLLAPTAYSQQRGLIGCIT	180		
DB	260 WGAIDTAACGDIISGLPVSAARGREIILGPANFEGCGRRLLAPTAYSQQRGLIGCIT	319		
QY	181 SLTGRDKNQVGEVQVWSTATQSEFLATCVNGVCTVFEHAGSKTLAGEKGEITQMYTNVD	240		
DB	320 SLTGRDKNQVGEVQVWSTATQSEFLATCVNGVCTVFEHAGSKTLAGEKGEITQMYTNVD	379		
QY	241 QDLVGMQAPPARSKNTPTCCSSDLVYTRHADVIPIPRRCDSGSLISPPVSYLKSS	300		
DB	380 QDLVGMQAPPARSKNTPTCCSSDLVYTRHADVIPIPRRCDSGSLISPPVSYLKSS	439		
QY	301 GGPLICPSGHAVGIFRAVCTRGAKXVDIFPVSMTETWR	341		
DB	440 GGPLICPSGHAVGIFRAVCTRGAKXVDIFPVSMTETWR	480		
RESULT 14				
068788				
AC	068788	PRELIMINARY;	PRT;	3010 AA.
DT	01-NOV-1996 (TREMBlrel. 01, Created)			
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	HCV polyprotein (Genome polypeptide).			
OS	Hepatitis C virus.			
CC	Virusess; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
CC	Hepadnavirinae;			
CC	NCBI_taxid=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96362158; PubMed=8720135;			
RA	Seki M., Honda Y.;			
RT	"Phosphorothioate antisense oligodeoxynucleotides capable of			
RT	inhibiting Hepatitis C virus gene expression: In vitro translation			
RT	assay."			
RL	J. Biochem. 118:1199-1204(1995).			
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A			
CC	LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:			
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF			
CC	PROTEIN C AND RNA (BY SIMILARITY).			
CC	EMBL, D45172, BA08120.1, -.			
DR	PIR; A61196; A61196.			
DR	PIR; P00246; P00246.			
DR	PIR; P00329; P00329.			
DR	HSSP; P26663; IUXP.			
DR	GO; GO:0016021; C: integral to membrane; IEA.			
DR	GO; GO:0019028; C: viral capsid; IEA.			
DR	GO; GO:0019031; C: viral envelope; IEA.			
DR	GO; GO:0005524; F: ATP binding; IEA.			
DR	GO; GO:0008026; F: ATP dependent helicase activity; IEA.			
DR	GO; GO:0003723; F: RNA binding; IEA.			
DR	GO; GO:0003568; F: RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008236; F: serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F: structural molecule activity; IEA.			
DR	GO; GO:0016740; F: transferase activity; IEA.			
DR	GO; GO:0006508; P: proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006350; P: transcription; IEA.			
DR	GO; GO:0019079; P: viral genome replication; IEA.			

DR GO:0019087; P:Viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002533; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_Rdrp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004103; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00398; Viral\_Rdrp; 1.  
 DR Pfam: PF0186062; HCV\_NS1; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEAD; 1.  
 DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; Transferrase; Transmembrane.  
 KW Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326880 MW; EED840E6A050E766 CRC64;

Query Match 96.2%; Score 1711; DB 12; Length 3010;  
 Best Local Similarity 94.7%; Pred. No. 4.6e-137;  
 Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

DR 1 VRGGRDAIILITCAVHELIIFDITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLV 60  
 DB 866 VRGGRDAIILITCAVHELIIFDITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLV 925  
 QY 61 RKAAGHYVOMAFKMAALGTCTYDHLTFLQDMANGLFDLAVANEPVFESMEYKIT 120  
 DB 926 WKAAGHYVOMAFKMAALGTCTYDHLTFLQDMANGLFDLAVANEPVFESMEYKIT 985  
 QY 121 WGAUTTAACGDIISGLPYSAARGREIILGPADNFEQGMRLAPITAYSOOTRGLGCIIT 180  
 DB 986 WGAUTTAACGDIISGLPYSAARGREIILGPADNFEQGMRLAPITAYSOOTRGLGCIIT 1045  
 QY 181 SLTGRDNQVGEVQVAVSTATQSFATCGVCTVTHGAGSKTLAPKPIITQMTTND 240  
 DB 1046 SLTGRDNQVGEVQVAVSTATQSFATCGVCTVTHGAGSKTLAPKPIITQMTTND 1105  
 QY 241 ODLYGMQAPGARSMTCCTGSSDLYLVTRHADVIPIRRRGDSRGLSPVSYLKSS 300  
 DB 1106 ODLYGMQAPGARSMTCCTGSSDLYLVTRHADVIPIRRRGDSRGLSPVSYLKSS 1165  
 QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDPIPVESMETTWK 341  
 DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDPIPVESMETTWK 1206  
 RESULT 15  
 P89966 PRELIMINARY; PRT; 3010 AA.  
 AC P89966;  
 DT 01-MAY-1997 (Tremblrel. 03. Created)  
 DT 01-MAY-1997 (Tremblrel. 03. Last annotation update)  
 DT 01-OCT-2003 (Tremblrel. 25. Last annotation update)  
 DE RNA for polyprotein (Genome polyprotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-type 1b;  
 RA Tanaka T.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-type 1b;  
 RA TANAKA T.;  
 RT "TORP";  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MENA (BY SIMILARITY).  
 CC EMBL: D89872; BAA14035.1; -.  
 DR PIR: A61196; A61196.  
 DR PIR: P0246; P0246.  
 DR PIR: P0804; P0804.  
 DR PIR: P0329; P0329.  
 DR HSP: P26653; IUXP.  
 DR GO:0016021; C:Integral to membrane; IEA.  
 DR GO:0019028; C:Viral capsid; IEA.  
 DR GO:0019031; C:Viral envelope; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO:0003723; F:RNA binding; IEA.  
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO:0005198; F:structural molecule activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO:0006350; P:transcription; IEA.  
 DR GO:0019079; P:viral genome replication; IEA.  
 DR GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR009003; Cys\_Ser\_lypsin.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002533; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_Rdrp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004103; Peptidase\_C29.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVlr.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00398; Viral\_Rdrp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEAD; 1.  
 KW CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; Transferrase; Transmembrane.  
 KW Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327023 MW; E075BD9CFDSD1261 CRC64;

Query Match 96.2%; Score 1711; DB 12; Length 3010;  
 Best Local Similarity 95.3%; Pred. No. 4.6e-137;  
 Matches 325; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

1 VRGGRDAIILITCAVHELIIFDITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLV 60

```

Db      866 VRGGRDAIILLTCVHPBELIFDIKLLALITGPIMLVLAQWTRVFPYFVRAQGLIRACMLV 925
QY      61  RKAAGHYVQMAFMKLAALTSTVYDHLTPLOMAHAGLRDLAVAVEPIFSMEVKIIT 120
Db      926 RKAAGHYVQMAFMKLAALTSTVYDHLTPLOMAHAGLRDLAVAVEPIFSMEVKIIT 985
QY      121 WGAADTAACGDIISGLPVSARRGREILGPADNFEQGMRLIAPITAYSOOTRGILGCIIT 180
Db      986 WGAADTAACGDIISGLPVSARRGREILGPADNFEQGMRLIAPITAYSOOTRGILGCIIT 1045
QY      181 SLTGRDNQVGEVQVAVSTATQSFATCVNGVCWTFHAGSKTLAGPKGPIITQMTNV 240
Db      1046 SLTGRDNQVGEVQVAVSTATQSFATCVNGVCWTFHAGSKTLAGPKGPIITQMTNV 1105
QY      241 QDLVGMQAPPGARSMPTCTGSSDLVLTNRHADVIPVRRGDSRGSLLSPRPVSYLKSS 300
Db      1106 QDLVGMQAPPGARSMPTCTGSSDLVLTNRHADVIPVRRGDSRGSLLSPRPVSYLKSS 1165
QY      301 GGPILCPSGHAWGIFRAAVCTRGVAKAVDFIPVESMETNR 341
Db      1166 GGPILCPSGHAWGIFRAAVCTRGVAKAVDFIPVESMETNR 1206

```

Search completed: May 6, 2004, 09:35:49  
 Job time : 32.8931 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.3837 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

Title: US-10-650-585-14

Perfect score: 1778  
Sequence: 1 VRGGRDAIILTCVHPELI.....RGVAKAVDFIVESMETTR 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625871 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PTCUTS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1702	95.7	2201	4	US-09-539-601-6
2	1702	95.7	2201	4	US-09-539-601-15
3	1702	95.7	3010	4	US-09-539-601-3
4	1702	95.7	3010	4	US-09-539-601-21
5	1702	95.7	3010	4	US-09-539-601-27
6	1698	95.5	1692	3	US-09-263-933-4
7	1698	95.5	1692	3	US-09-263-933-2
8	1698	95.5	2307	3	US-09-263-933-1
9	1698	95.5	2307	3	US-09-263-933-11
10	1695	95.3	1692	3	US-09-263-933-9
11	1695	95.3	1692	3	US-09-263-933-9
12	1695	95.3	2307	3	US-09-263-933-9
13	1695	95.3	2307	3	US-09-263-933-9
14	1695	95.3	3010	4	US-09-539-601-33
15	1687	94.9	3010	3	US-09-014-416-3
16	1686	94.8	1692	3	US-09-263-933-18
17	1686	94.8	1692	3	US-09-263-933-18
18	1686	94.8	2307	3	US-09-263-933-16
19	1686	94.8	2307	3	US-09-263-933-16
20	1649	92.7	2013	1	US-08-324-977-12
21	1649	92.7	2013	2	US-08-384-616-12
22	1649	92.7	2013	2	US-08-904-686A-12
23	1649	92.7	2013	3	US-09-315-850-12
24	1649	92.7	2201	4	US-08-953-981A-2
25	1649	92.7	2620	1	US-08-324-977-32
26	1649	92.7	2620	2	US-08-384-616-32
27	1649	92.7	2620	2	US-08-904-686A-32

28	1649	92.7	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1649	92.7	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1649	92.7	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1649	92.7	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1649	92.7	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1649	92.7	3010	1	US-08-324-977-2	Sequence 2, Appl
34	1649	92.7	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1649	92.7	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1649	92.7	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1649	92.7	3010	2	US-08-904-686A-2	Sequence 2, Appl
38	1649	92.7	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1649	92.7	3010	3	US-09-315-850-12	Sequence 12, Appl
40	1649	92.7	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1570	88.3	1648	1	US-08-188-281B-12	Sequence 12, Appl
42	1570	88.3	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1570	88.3	1648	5	PCT-US95-01087-12	Sequence 12, Appl
44	1570	88.3	3011	1	US-08-188-281B-1	Sequence 1, Appl
45	1570	88.3	3011	1	US-08-453-552-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-539-601-6  
Sequence 6, Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Bartschlag, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 6  
LENGTH: 2201  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-6

Query Match	95.7%;	Score 1702;	DB 4;	Length 2201;		
Best Local Similarity	94.1%;	Pred. No. 4e-164;				
Matches 321;	Conservative 11;	Mismatches 9;	Indels 0;	Gaps 0		
QY	1	VRGGRDAIILTCVHPELI	FDITKLLAI	FGPLMVLQGITKVPYFVRAQGLIRACMLV	60	
Db	57	VRGGRDAVILTCVHPELI	FDITKLLAI	FGPLMVLQGITKVPYFVRAQGLIRACMLV	116	
QY	61	RKAAGHYQMAFMKLAAL	TGYVVDHLP	LDQMAVAGRDIAVAEPIFSDMEVKIT	120	
Db	117	RKAAGHYQMAFMKLAAL	TGYVVDHLP	LDQMAVAGRDIAVAEPIFSDMEVKIT	176	
QY	121	WGADTACDIIISGLPVS	ARRGRIIL	GPADNFEQGMWLLAPITAYSCQTRGLGCIIT	180	
Db	177	WGADTACDIIISGLPVS	ARRGRIIL	GPADNFEQGMWLLAPITAYSCQTRGLGCIIT	236	
QY	181	SILTRGRKNOVEGVQV	STATOSFLAT	CNGICWYFHAGSKTLA	PGPIITQWYTNVD	240
Db	237	SILTRGRKNOVEGVQV	STATOSFLAT	CNGICWYFHAGSKTLA	PGPIITQWYTNVD	296
QY	241	ODIVGQAQPGARSMPT	CGSSDLYL	VTIRADVIVP	RRGDSRGLSPREVSYLKSS	300
Db	297	ODIVGQAQPGARSMPT	CGSSDLYL	VTIRADVIVP	RRGDSRGLSPREVSYLKSS	356
QY	301	GGPLCPSGHNAVIGIP	RAVCTRGVAKAVD	FIVESMETTR	341	
Db	357	GGPLCPSGHNAVIGIP	RAVCTRGVAKAVD	FIVESMETTR	397	

RESULT 2



RESULT 5  
US-09-539-601-27  
; Sequence 27, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Barenshlager, Ralf FM  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; EARLIER FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-27

Query Match 95.7%; Score 1702; DB 4; Length 3010;  
Best Local Similarity 94.1%; Pred. No. 6.3e-164;  
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 60  
DB VRGGRDAVILITCAIHEPILFTITKLLALIGPLWVLOAGITVFPVFAOGLIRACMLV 925  
QY 61 RKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 120  
DB RKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 985  
QY 926 RKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 985  
DB RKAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 985  
QY 121 WGAADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 180  
DB WGAADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 1045  
QY 986 WGAADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 1045  
DB WGAADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 1045  
QY 181 SLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 240  
DB SLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 1105  
QY 1046 SLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 1105  
DB SLGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 1105  
QY 241 QDLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 300  
DB QDLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 300  
QY 1106 QDLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 1165  
DB QDLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 1165  
QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTR 341  
DB GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTR 1206

RESULT 6  
US-09-263-933-4  
; Sequence 4, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/263,933  
; EARLIER FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: 09/129,611  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
US-09-263-933-4

Query Match 95.5%; Score 1698; DB 3; Length 1692;

Best Local Similarity 94.1%; Pred. No. 6.3e-164;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 61  
DB RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 205  
QY 146 RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 205  
DB RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 205  
QY 62 KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 121  
DB KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 265  
QY 206 KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 265  
DB KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 265  
QY 122 GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 181  
DB GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 325  
QY 266 GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 325  
DB GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 325  
QY 182 LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 241  
DB LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 385  
QY 326 LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 385  
DB LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 385  
QY 242 DLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 301  
DB DLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 445  
QY 386 DLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 445  
DB DLVGMAPPGARSMPTCTGSSDLYLVTRHADVIPIRRRGDSRGLSPRPVSYLKSSG 445  
QY 302 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTR 341  
DB GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVSMETTR 465

RESULT 7  
US-09-919-901-4  
; Sequence 4, Application US/09919901  
; Patent No. 6539738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; EARLIER FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263,933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129,611  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
US-09-919-901-4

Query Match 95.5%; Score 1698; DB 4; Length 1692;  
Best Local Similarity 94.1%; Pred. No. 6.3e-164;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 61  
DB RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 205  
QY 146 RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 205  
DB RGRDAIILITCAVHEPILFDITKLLAIFGPIWVLOAGITVFPVFAOGLIRACMLV 205  
QY 62 KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 121  
DB KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 265  
QY 206 KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 265  
DB KAAGHYVQMAFMKLAALGTIVYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIIT 265  
QY 122 GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 181  
DB GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 325  
QY 266 GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 325  
DB GADTAACGDIISGLPVSAARGREIILGPADNLEGQGMRLAPITAYSOQTRGLGCIIT 325  
QY 182 LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 241  
DB LTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPKPITOMYTND 241

Db 326 LTGRDKNQVEGEVQVSTATQSFATCVCNGCWTYHAGSKTLAAGPKBITOMTYNVQ 385  
Qy 242 DLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSGSLSPRPVSYLKSSG 301  
Db 386 DLVGMQAPPGARSLTPTCGSSDLYLVTRHADVIPIVRRGDSGSLSPRPVSYLKSSG 445  
Qy 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 341  
Db 446 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 485

RESULT 8  
US-09-263-933-2  
Sequence 2, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
EARLIER FILING DATE: 1999-03-08  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-2

Query Match 95.5%; Score 1698; DB 3; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 1.1e-163;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGRDAILLTCVAHPFLIFDITKLLAIFGPIMLVLOAGITKYFVRAQGLIRACMLVR 61  
Db 238 RGRDAILLTCVAHPFLIFDITKLLAIFGPIMLVLOAGITKYFVRAQGLIRACMLVR 297  
Qy 62 KAAGHYVQAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPIFSDEMKIITW 121  
Db 298 KAAGHYVQAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPIFSDEMKIITW 357  
Qy 122 GADTAACGDIILGIPVSARRGKEILLGPADNFEQGRLLAPIYASQOIRGLIGCIITS 181  
Db 358 GADTAACGDIILGIPVSARRGKEILLGPADNFEQGRLLAPIYASQOIRGLIGCIITS 417  
Qy 182 LTGRDKNQVEGEVQVSTATQSFATCVCNGCWTYHAGSKTLAAGPKBITOMTYNVQ 241  
Db 418 LTGRDKNQVEGEVQVSTATQSFATCVCNGCWTYHAGSKTLAAGPKBITOMTYNVQ 477  
Qy 242 DLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSGSLSPRPVSYLKSSG 301  
Db 478 DLVGMQAPPGARSLTPTCGSSDLYLVTRHADVIPIVRRGDSGSLSPRPVSYLKSSG 537  
Qy 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 341  
Db 538 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 577

RESULT 9  
US-09-919-901-2  
Sequence 2, Application US/09919901  
Patent No. 6599738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919,901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-2

Query Match 95.5%; Score 1698; DB 4; Length 2307;  
Best Local Similarity 94.1%; Pred. No. 1.1e-163;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGRDAILLTCVAHPFLIFDITKLLAIFGPIMLVLOAGITKYFVRAQGLIRACMLVR 61  
Db 238 RGRDAILLTCVAHPFLIFDITKLLAIFGPIMLVLOAGITKYFVRAQGLIRACMLVR 297  
Qy 62 KAAGHYVQAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPIFSDEMKIITW 121  
Db 298 KAAGHYVQAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPIFSDEMKIITW 357  
Qy 122 GADTAACGDIILGIPVSARRGKEILLGPADNFEQGRLLAPIYASQOIRGLIGCIITS 181  
Db 358 GADTAACGDIILGIPVSARRGKEILLGPADNFEQGRLLAPIYASQOIRGLIGCIITS 417  
Qy 182 LTGRDKNQVEGEVQVSTATQSFATCVCNGCWTYHAGSKTLAAGPKBITOMTYNVQ 241  
Db 418 LTGRDKNQVEGEVQVSTATQSFATCVCNGCWTYHAGSKTLAAGPKBITOMTYNVQ 477  
Qy 242 DLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVIPIVRRGDSGSLSPRPVSYLKSSG 301  
Db 478 DLVGMQAPPGARSLTPTCGSSDLYLVTRHADVIPIVRRGDSGSLSPRPVSYLKSSG 537  
Qy 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 341  
Db 538 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 577

RESULT 10  
US-09-263-933-11  
Sequence 11, Application US/09263933  
Patent No. 6280940  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/263,933  
EARLIER FILING DATE: 1999-03-08  
EARLIER FILING DATE: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
US-09-263-933-11

Query Match 95.3%; Score 1695; DB 3; Length 1692;  
Best Local Similarity 93.8%; Pred. No. 1.4e-153;  
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

```

QY 2 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLAQGITKYPVFRAGGILRACMLVR 61
DB 146 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLAQGITKYPVFRAGGILRACMLVR 205
QY 62 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEKILITM 121
DB 206 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEKILITM 265
QY 122 GADTAACGDIISGLPVASARGREIILGPADNFBEGGWRLLAPITAYSQOTRGLGCIITIS 181
DB 266 GADTAACGDIISGLPVASARGREIILGPADNFBEGGWRLLAPITAYSQOTRGLGCIITIS 325
QY 182 LTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 241
DB 326 LTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 385
QY 242 DLVGMQAPPGARSMTPTCTGSSDLVLTVRHADVIPIVRRGDSRGLSPRPVSYLKSSG 301
DB 386 DLVGMQAPPGARSMTPTCTGSSDLVLTVRHADVIPIVRRGDSRGLSPRPVSYLKSSG 445
QY 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 341
DB 446 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 485

RESULT 11
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-11

Query Match          95.3%; Score 1695; DB 4; Length 1692;
Best Local Similarity 93.8%; Pred. No. 1.4e-163;
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

```

```

DB 386 DLVGMQAPPGARSMTPTCTGSSDLVLTVRHADVIPIVRRGDSRGLSPRPVSYLKSSG 445
QY 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 341
DB 446 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 485

RESULT 12
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

Query Match          95.3%; Score 1695; DB 3; Length 2307;
Best Local Similarity 93.8%; Pred. No. 2.2e-163;
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLAQGITKYPVFRAGGILRACMLVR 61
DB 238 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLAQGITKYPVFRAGGILRACMLVR 297
QY 62 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEKILITM 121
DB 298 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAVEPVFSDEMEKILITM 357
QY 122 GADTAACGDIISGLPVASARGREIILGPADNFBEGGWRLLAPITAYSQOTRGLGCIITIS 181
DB 358 GADTAACGDIISGLPVASARGREIILGPADNFBEGGWRLLAPITAYSQOTRGLGCIITIS 417
QY 182 LTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 241
DB 418 LTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 477
QY 242 DLVGMQAPPGARSMTPTCTGSSDLVLTVRHADVIPIVRRGDSRGLSPRPVSYLKSSG 301
DB 478 DLVGMQAPPGARSMTPTCTGSSDLVLTVRHADVIPIVRRGDSRGLSPRPVSYLKSSG 537
QY 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 341
DB 538 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTMR 577

RESULT 13
US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02

```

PRIOR APPLICATION NUMBER: 09/263,933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129,611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: :  
US-09-919-901-9

Query Match 95.3%; Score 1695; DB 4; Length 2307;  
Best Local Similarity 93.8%; Pred. No. 2.2e-163;  
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGGSDAIIITLTCVHPELIFDITKLLATFGPLMVLQAGITKPYFVRAGLIRACMLV 61  
DB 238 RGGSDAIIITLTCVHPELIFDITKLLATFGPLMVLQAGITKPYFVRAGLIRACMLV 297  
QY 62 KAAAGHYVQMAFMKLAALGTYYVDLTPQDMAHAGRLDLAVAVEPVIFSDMEVKIITW 121  
DB 298 KAAAGHYVQMAFMKLAALGTYYVDLTPQDMAHAGRLDLAVAVEPVIFSDMEVKIITW 357  
QY 122 GADTAAAGDIIISGLPVSARRGREILGPADNFGGCRLLAPITTAASQOTRGLGCIIT 161  
DB 358 GADTAAAGDIIISGLPVSARRGREILGPADNFGGCRLLAPITTAASQOTRGLGCIIT 417  
QY 182 LTGSDKNQVGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAPKPGITOMYTNVD 241  
DB 418 LTGSDKNQVGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAPKPGITOMYTNVD 477  
QY 242 DLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKGSS 301  
DB 478 DLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKGSS 537  
QY 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341  
DB 538 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 577

RESULT 14  
US-09-539-601-33  
Sequence 33; Application US/09539601C  
Patent No. 6630343  
GENERAL INFORMATION:  
APPLICANT: Barenscnager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539,601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 33  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-539-601-33

Query Match 95.3%; Score 1695; DB 4; Length 3010;  
Best Local Similarity 93.8%; Pred. No. 3.3e-163;  
Matches 320; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 VRGRDIIITLTCVHPELIFDITKLLATFGPLMVLQAGITKPYFVRAGLIRACMLV 60  
DB 866 VRGRDIIITLTCVHPELIFDITKLLATFGPLMVLQAGITKPYFVRAGLIRACMLV 925  
QY 61 RKAAGHYVQMAFMKLAALGTYYVDLTPQDMAHAGRLDLAVAVEPVIFSDMEVKIIT 120

DB 926 RKAAGHYVQMAFMKLAALGTYYVDLTPQDMAHAGRLDLAVAVEPVIFSDMEVKIIT 985  
QY 121 WADTAAAGDIIISGLPVSARRGREILGPADNFGGCRLLAPITTAASQOTRGLGCIIT 180  
DB 986 WADTAAAGDIIISGLPVSARRGREILGPADNFGGCRLLAPITTAASQOTRGLGCIIT 1045  
QY 181 SLTGRDNQVGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAPKPGITOMYTNVD 240  
DB 1046 SLTGRDNQVGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAPKPGITOMYTNVD 1105  
QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKGSS 300  
DB 1106 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKGSS 1165  
QY 301 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341  
DB 1166 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 15  
US-09-014-416-3  
Sequence 3; Application US/09014416  
Patent No. 6153421  
GENERAL INFORMATION:  
APPLICANT: Yanagi, Masayuki  
APPLICANT: Buht, Jens  
APPLICANT: Emerson, Susanne U.  
APPLICANT: Purcell, Robert H.  
TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND  
FILE REFERENCE: 20264276  
CURRENT APPLICATION NUMBER: US/09/014,416  
CURRENT FILING DATE: 1998-01-27  
EARLIER APPLICATION NUMBER: US 60/053,062  
EARLIER FILING DATE: 1997-07-18  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-014-416-3

Query Match 94.9%; Score 1687; DB 3; Length 3010;  
Best Local Similarity 93.3%; Pred. No. 2.2e-162;  
Matches 318; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 VRGRDIIITLTCVHPELIFDITKLLATFGPLMVLQAGITKPYFVRAGLIRACMLV 60  
DB 866 VRGRDIIITLTCVHPELIFDITKLLATFGPLMVLQAGITKPYFVRAGLIRACMLV 925  
QY 61 RKAAGHYVQMAFMKLAALGTYYVDLTPQDMAHAGRLDLAVAVEPVIFSDMEVKIIT 120  
DB 926 RKAAGHYVQMAFMKLAALGTYYVDLTPQDMAHAGRLDLAVAVEPVIFSDMEVKIIT 985  
QY 121 WADTAAAGDIIISGLPVSARRGREILGPADNFGGCRLLAPITTAASQOTRGLGCIIT 180  
DB 986 WADTAAAGDIIISGLPVSARRGREILGPADNFGGCRLLAPITTAASQOTRGLGCIIT 1045  
QY 181 SLTGRDNQVGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAPKPGITOMYTNVD 240  
DB 1046 SLTGRDNQVGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAPKPGITOMYTNVD 1105  
QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKGSS 300  
DB 1106 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKGSS 1165  
QY 301 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341  
DB 1166 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

Search completed: May 6, 2004, 09:39:05

Fri May 7 13:37:05 2004

Job time : 14.3837 secs

us-10-650-585-14.ra1

Page 7





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 34.8831 Seconds  
(without alignments)  
2713.357 Million cell updates/sec

Title: US-10-650-585-14

Perfect score: 1778  
Sequence: 1 VRGGRDAIILLTCVAHPBELI.....RGVAKAVDFIPVBSMETTWK 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapexc 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCRUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1778	100.0	341	US-10-017-736-14	Sequence 14, Appl1
2	1778	100.0	341	US-10-650-585-14	Sequence 14, Appl1
3	1778	100.0	352	US-10-017-736-13	Sequence 13, Appl1
4	1778	100.0	352	US-10-650-585-13	Sequence 13, Appl1
5	1778	100.0	380	US-10-017-736-12	Sequence 12, Appl1
6	1778	100.0	380	US-10-650-585-12	Sequence 12, Appl1
7	1778	100.0	393	US-10-017-736-11	Sequence 11, Appl1
8	1778	100.0	393	US-10-650-585-11	Sequence 11, Appl1
9	1778	100.0	403	US-10-017-736-2	Sequence 2, Appl1
10	1778	100.0	409	US-10-650-585-2	Sequence 2, Appl1
11	1702	95.7	2201	US-10-029-907-3	Sequence 3, Appl1
12	1702	95.7	2201	US-10-309-561-3	Sequence 3, Appl1
13	1702	95.7	3010	US-10-467-000-1	Sequence 1, Appl1
14	1698	95.5	1692	US-09-919-901-4	Sequence 4, Appl1
15	1698	95.5	1692	US-10-191-966-4	Sequence 4, Appl1

16	1698	95.5	2307	10	US-09-919-901-2	Sequence 2, Appl1
17	1698	95.5	2307	14	US-10-191-966-2	Sequence 2, Appl1
18	1695	95.3	1692	14	US-09-919-901-11	Sequence 11, Appl1
19	1695	95.3	1692	14	US-10-191-966-11	Sequence 11, Appl1
20	1695	95.3	2307	14	US-09-919-901-9	Sequence 9, Appl1
21	1695	95.3	2307	14	US-10-191-966-9	Sequence 9, Appl1
22	1686	94.8	1692	14	US-09-919-901-18	Sequence 18, Appl1
23	1686	94.8	1692	14	US-10-191-966-18	Sequence 18, Appl1
24	1686	94.8	2307	14	US-09-919-901-16	Sequence 16, Appl1
25	1686	94.8	2307	14	US-10-191-966-16	Sequence 16, Appl1
26	1649	92.7	2201	13	US-10-085-476-2	Sequence 2, Appl1
27	1589	89.4	303	13	US-10-017-736-10	Sequence 10, Appl1
28	1589	89.4	303	16	US-10-650-585-10	Sequence 10, Appl1
29	1589	89.4	334	13	US-10-017-726-4	Sequence 4, Appl1
30	1589	89.4	334	16	US-10-650-585-4	Sequence 4, Appl1
31	1580	88.9	303	13	US-10-017-736-18	Sequence 18, Appl1
32	1580	88.9	303	16	US-10-650-585-18	Sequence 18, Appl1
33	1579	88.8	303	13	US-10-017-736-16	Sequence 16, Appl1
34	1579	88.8	303	16	US-10-650-585-16	Sequence 16, Appl1
35	1570	88.3	301	13	US-10-017-736-17	Sequence 17, Appl1
36	1570	88.3	301	16	US-10-650-585-17	Sequence 17, Appl1
37	1569	88.2	3011	9	US-09-742-653-4	Sequence 4, Appl1
38	1569	88.2	3011	10	US-09-891-894-3	Sequence 3, Appl1
39	1569	88.2	3011	14	US-10-184-150-3	Sequence 3, Appl1
40	1569	88.2	3011	15	US-10-328-997-3	Sequence 3, Appl1
41	1569	88.2	3012	9	US-09-238-076-2	Sequence 2, Appl1
42	1569	88.2	3012	10	US-09-995-937-2	Sequence 2, Appl1
43	1569	88.1	3012	10	US-09-917-563-2	Sequence 2, Appl1
44	1566	88.1	3011	12	US-10-296-734-406	Sequence 406, App
45	1565	88.0	3011	9	US-09-916-353-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-10-017-736-14 Application US/10017736  
; Sequence 14, US-10-017-736-14  
; Publication No. US20020192640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/017,736  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256,031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: HCV  
US-10-017-736-14

Query Match	100.0%	Score 1778;	DB 13;	Length 341;
Best Local Similarity	100.0%	Pred. No. 2.7e-174;	Indels 0;	Gaps 0;
Matches 341;	Conservative 0;	Mismatches 0;		
QY	1	VRGGRDAIILLTCVAHPBELI	FDITKLLAIFGFLAVLQGITKVFYFVAAGIIRACMLV	60
DB	1	VRGGRDAIILLTCVAHPBELI	FDITKLLAIFGFLAVLQGITKVFYFVAAGIIRACMLV	60
QY	61	RKAAGHYVQMAFMKLAALGTIVYDHLTPIQDWAHGRDLAAVEPVI	PSDMEYKIIT	120
DB	61	RKAAGHYVQMAFMKLAALGTIVYDHLTPIQDWAHGRDLAAVEPVI	PSDMEYKIIT	120
QY	121	WGADTLAAGDITISGLFVSARRGREILLGPADNEGQWMLLPITV	SOOTGLGCIIT	180
DB	121	WGADTLAAGDITISGLFVSARRGREILLGPADNEGQWMLLPITV	SOOTGLGCIIT	180
QY	181	SLTGRDKQVGEQVAVSTATOSFIATCVNGVCMVFHAGSKTLAGP	KPITOMTVND	240
DB	181	SLTGRDKQVGEQVAVSTATOSFIATCVNGVCMVFHAGSKTLAGP	KPITOMTVND	240

Db 181 SLTGRDNQVEGEVQVSTATQSFATCNGVCMTVFHGAAGSKTLAAGPKGITQMTYND 240  
 QY 241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRVSYLKSS 300  
 Db 241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRVSYLKSS 300  
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 Db 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341

## RESULT 2

US-10-650-585-14  
 ; Sequence 14, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14  
 ; LENGTH: 341  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-14

Query Match 100.0%; Score 1778; DB 16; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-174;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60  
 Db 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60  
 QY 61 RKAAGHYVQMAFPMKLAALGTGYVDHLTPLOMAHAGLBDLAAVEPVFSDMEVKIIT 120  
 Db 61 RKAAGHYVQMAFPMKLAALGTGYVDHLTPLOMAHAGLBDLAAVEPVFSDMEVKIIT 120  
 QY 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180  
 Db 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180  
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCNGVCMTVFHGAAGSKTLAAGPKGITQMTYND 240  
 Db 181 SLTGRDNQVEGEVQVSTATQSFATCNGVCMTVFHGAAGSKTLAAGPKGITQMTYND 240  
 QY 241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRVSYLKSS 300  
 Db 241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRVSYLKSS 300  
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 Db 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341

## RESULT 3

US-10-017-736-13  
 ; Sequence 13, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-017-736-13

Query Match 100.0%; Score 1778; DB 13; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-174;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60  
 Db 12 VRGGRDAIILITCAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 71  
 QY 61 RKAAGHYVQMAFPMKLAALGTGYVDHLTPLOMAHAGLBDLAAVEPVFSDMEVKIIT 120  
 Db 72 RKAAGHYVQMAFPMKLAALGTGYVDHLTPLOMAHAGLBDLAAVEPVFSDMEVKIIT 131  
 QY 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180  
 Db 132 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 191  
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCNGVCMTVFHGAAGSKTLAAGPKGITQMTYND 240  
 Db 192 SLTGRDNQVEGEVQVSTATQSFATCNGVCMTVFHGAAGSKTLAAGPKGITQMTYND 251  
 QY 241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRVSYLKSS 300  
 Db 252 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRVSYLKSS 311  
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341  
 Db 312 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 352

## RESULT 4

US-10-650-585-13  
 ; Sequence 13, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650,585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017,736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-13

Query Match 100.0%; Score 1778; DB 16; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-174;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60  
 Db 12 VRGGRDAIILITCAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 71  
 QY 61 RKAAGHYVQMAFPMKLAALGTGYVDHLTPLOMAHAGLBDLAAVEPVFSDMEVKIIT 120  
 Db 72 RKAAGHYVQMAFPMKLAALGTGYVDHLTPLOMAHAGLBDLAAVEPVFSDMEVKIIT 131  
 QY 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180

Fri May 7 13:37:05 2004

us-10-650-585-14.rapb

Page 3

```
Db 132 WGADTAACGDIISGLPVSARGREILLGPADNFEQGWRLAPITAYSOQTRGLGCIIT 191
Qy 181 SLTGRDNQVGEVGVVSTATQSFATCNVGVCMVTFHAGSKTLAGPKGPIITQMTYND 240
Db 192 SLTGRDNQVGEVGVVSTATQSFATCNVGVCMVTFHAGSKTLAGPKGPIITQMTYND 251
Qy 241 QDLVGMQAPPGARSMTPTCGSSDLYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300
Db 252 QDLVGMQAPPGARSMTPTCGSSDLYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 311
Qy 301 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db 312 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 352
```

```
RESULT 5
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12
```

```
Query Match 100.0%; Score 1778; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 3,2e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VRGSDAIIILTCVAHPDLIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60
Db 40 VRGSDAIIILTCVAHPDLIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 99
Qy 61 RKAAGHYVQMAFMKLAALTGTIVYDHLTPLODMAHAGRLDAVAVEPIFSDEVKIIT 120
Db 100 RKAAGHYVQMAFMKLAALTGTIVYDHLTPLODMAHAGRLDAVAVEPIFSDEVKIIT 159
Qy 121 WGADTAACGDIISGLPVSARGREILLGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db 160 WGADTAACGDIISGLPVSARGREILLGPADNFEQGWRLAPITAYSOQTRGLGCIIT 219
Qy 181 SLTGRDNQVGEVGVVSTATQSFATCNVGVCMVTFHAGSKTLAGPKGPIITQMTYND 240
Db 220 SLTGRDNQVGEVGVVSTATQSFATCNVGVCMVTFHAGSKTLAGPKGPIITQMTYND 279
Qy 241 QDLVGMQAPPGARSMTPTCGSSDLYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300
Db 280 QDLVGMQAPPGARSMTPTCGSSDLYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 339
Qy 301 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db 340 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 380
```

```
RESULT 6
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
```

```
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-12
```

```
Query Match 100.0%; Score 1778; DB 16; Length 380;
Best Local Similarity 100.0%; Pred. No. 3,2e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VRGSDAIIILTCVAHPDLIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60
Db 40 VRGSDAIIILTCVAHPDLIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 99
Qy 61 RKAAGHYVQMAFMKLAALTGTIVYDHLTPLODMAHAGRLDAVAVEPIFSDEVKIIT 120
Db 100 RKAAGHYVQMAFMKLAALTGTIVYDHLTPLODMAHAGRLDAVAVEPIFSDEVKIIT 159
Qy 121 WGADTAACGDIISGLPVSARGREILLGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db 160 WGADTAACGDIISGLPVSARGREILLGPADNFEQGWRLAPITAYSOQTRGLGCIIT 219
Qy 181 SLTGRDNQVGEVGVVSTATQSFATCNVGVCMVTFHAGSKTLAGPKGPIITQMTYND 240
Db 220 SLTGRDNQVGEVGVVSTATQSFATCNVGVCMVTFHAGSKTLAGPKGPIITQMTYND 279
Qy 241 QDLVGMQAPPGARSMTPTCGSSDLYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300
Db 280 QDLVGMQAPPGARSMTPTCGSSDLYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 339
Qy 301 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db 340 GGFLPCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTMR 380
```

```
RESULT 7
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11
```

```
Query Match 100.0%; Score 1778; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 3,4e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VRGSDAIIILTCVAHPDLIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 60
Db 53 VRGSDAIIILTCVAHPDLIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACMLV 112
Qy 61 RKAAGHYVQMAFMKLAALTGTIVYDHLTPLODMAHAGRLDAVAVEPIFSDEVKIIT 120
```

```
Db      113 RKAAGHYVQMAFMKLAALGTGYVDHLFPLQDMAHAGRLDAVAEPIFSDMEVKIIT 172
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
; US-10-650-585-11

Query Match      100.0%; Score 1778; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.4e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VRGGRDAIILLTCVHPELIFDITKLLAIFGLPMVLQAGITKVPYFVRAQGLIRACMLV 60
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
; US-10-650-585-11

Db      53 VRGGRDAIILLTCVHPELIFDITKLLAIFGLPMVLQAGITKVPYFVRAQGLIRACMLV 112
QY      61 RKAAGHYVQMAFMKLAALGTGYVDHLFPLQDMAHAGRLDAVAEPIFSDMEVKIIT 120
Db      113 RKAAGHYVQMAFMKLAALGTGYVDHLFPLQDMAHAGRLDAVAEPIFSDMEVKIIT 172
QY      121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
Db      173 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 232
QY      181 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTFPHAGSKTLAGPKPIITQMYTND 240
Db      233 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTFPHAGSKTLAGPKPIITQMYTND 292
QY      241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLSPRPVSYLKSS 300
Db      293 QDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLSPRPVSYLKSS 352
QY      301 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db      353 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTMR 393
```

RESULT 9  
US-10-017-736-2  
Sequence 2, Application US/10017736

```
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
; US-10-017-736-2

Query Match      100.0%; Score 1778; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.6e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VRGGRDAIILLTCVHPELIFDITKLLAIFGLPMVLQAGITKVPYFVRAQGLIRACMLV 60
Db      57 VRGGRDAIILLTCVHPELIFDITKLLAIFGLPMVLQAGITKVPYFVRAQGLIRACMLV 116
QY      61 RKAAGHYVQMAFMKLAALGTGYVDHLFPLQDMAHAGRLDAVAEPIFSDMEVKIIT 120
Db      117 RKAAGHYVQMAFMKLAALGTGYVDHLFPLQDMAHAGRLDAVAEPIFSDMEVKIIT 176
QY      121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 236
QY      181 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTFPHAGSKTLAGPKPIITQMYTND 240
Db      237 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMTFPHAGSKTLAGPKPIITQMYTND 296
QY      241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLSPRPVSYLKSS 300
Db      297 QDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGLSPRPVSYLKSS 356
QY      301 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db      357 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397
```

```
RESULT 10
US-10-650-585-2
; Sequence 2, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
; US-10-650-585-2
```

Query Match 100.0%; Score 1778; DB 16; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.6e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILLTCVHPELIFDITKLLAIFGLPMVLQAGITKVPYFVRAQGLIRACMLV 60

```

Db      57 VRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPVSDEVEKIIIT 120
Db      117 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPVSDEVEKIIIT 176
Qy      121 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFVHGASKTLAGPKGITOMYTND 240
Db      237 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFVHGASKTLAGPKGITOMYTND 296
Qy      241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRPVSYLKSS 300
Db      297 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRPVSYLKSS 356
Qy      301 GGFLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db      357 GGFLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 397

```

RESULT 11

```

US-10-029-907-3
; Sequence 3, Application US/10029907
; Publication No. US20020142350A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
US-10-029-907-3

```

Query Match 95.7%; Score 1702; DB 13; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 2.6e-165;  
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

Qy      1 VRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 60
Db      57 VRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPVSDEVEKIIIT 120
Db      117 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPVSDEVEKIIIT 176
Qy      121 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFVHGASKTLAGPKGITOMYTND 240
Db      237 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFVHGASKTLAGPKGITOMYTND 296
Qy      241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRPVSYLKSS 300

```

```

Db      297 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRPVSYLKSS 356
Qy      301 GGFLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db      357 GGFLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 397

```

RESULT 12

```

US-10-309-561-3
; Sequence 3, Application US/10309561
; Publication No. US20030148348A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/309,561
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US/10/029,907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
US-10-309-561-3

```

Query Match 95.7%; Score 1702; DB 14; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 2.6e-165;  
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

Qy      1 VRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 60
Db      57 VRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPVSDEVEKIIIT 120
Db      117 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPVSDEVEKIIIT 176
Qy      121 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFVHGASKTLAGPKGITOMYTND 240
Db      237 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFVHGASKTLAGPKGITOMYTND 296
Qy      241 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRPVSYLKSS 300
Db      297 QDLVGMQAPPGARSMTPTCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPRPVSYLKSS 356
Qy      301 GGFLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db      357 GGFLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 397

```

```

; APPLICANT: De Francesco, Raffaele
; APPLICANT: Migliaccio, Giovanni
; APPLICANT: Paonessa, Giacomo
; TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
; TITLE OF INVENTION: ENHANCED CELLS
; FILE REFERENCE: ITRO003P
; CURRENT APPLICATION NUMBER: US/10/467,000
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: PCT/EP02/00526
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/263,479
; PRIOR FILING DATE: 2001-01-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Con 1 HCV isolate nucleic acid
; US-10-467-000-1

```

```

Query Match      95.7%; Score 1702; DB 12; Length 3010;
Best Local Similarity 94.1%; Pred. No. 4,1e-165;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 VGGSDAIIILTCVAHPELIPDITKLLAIFGPMVLOAGITKYPYFVRAQGLIRACMLV 60
DB 866 VGGSDAIIILTCVAHPELIPDITKLLAIFGPMVLOAGITKYPYFVRAQGLIRACMLV 925
QY 61 KKAAGHYQMAFMKLAALITGYVDHLTPLOMAHAGLRLAVALVEPVFSDMEVKIIT 120
DB 926 KKAAGHYQMAFMKLAALITGYVDHLTPLOMAHAGLRLAVALVEPVFSDMEVKIIT 985
QY 121 WGNADTAACGDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 180
DB 986 WGNADTAACGDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 1045
QY 181 SITGDRKQVEGEVGVSTATQSFPLATCVNGVCTVFGAGSKTLAPKGPITQMTTNVD 240
DB 1046 SITGDRKQVEGEVGVSTATQSFPLATCVNGVCTVFGAGSKTLAPKGPITQMTTNVD 1105
QY 241 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSGSLSPRPVSYLKGSSG 300
DB 1106 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSGSLSPRPVSYLKGSSG 1165
QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDPFVPSMETTWR 341
DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDPFVPSMETTWR 1206

```

```

RESULT 14
US-09-919-901-4
; Sequence 4, Application US/09919901
; Publication No. US20030082518A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION:
; US-09-919-901-4

```

```

Query Match      95.5%; Score 1698; DB 10; Length 1692;
Best Local Similarity 94.1%; Pred. No. 4.6e-165;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 RGRDAIILTCVAHPELIPDITKLLAIFGPMVLOAGITKYPYFVRAQGLIRACMLV 61
DB 146 RGRDAIILTCVAHPELIPDITKLLAIFGPMVLOAGITKYPYFVRAQGLIRACMLV 205
QY 62 KKAAGHYQMAFMKLAALITGYVDHLTPLOMAHAGLRLAVALVEPVFSDMEVKIIT 121
DB 206 KKAAGHYQMAFMKLAALITGYVDHLTPLOMAHAGLRLAVALVEPVFSDMEVKIIT 265
QY 122 GADTAACGDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 181
DB 266 GADTAACGDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 325
QY 182 LTGRPKQVEGEVGVSTATQSFPLATCVNGVCTVFGAGSKTLAPKGPITQMTTNVD 241
DB 326 LTGRPKQVEGEVGVSTATQSFPLATCVNGVCTVFGAGSKTLAPKGPITQMTTNVD 385
QY 242 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSGSLSPRPVSYLKGSSG 301
DB 386 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSGSLSPRPVSYLKGSSG 445
QY 302 GGPLCPGSHAVGIFRAVCTRGVAKAVDPFVPSMETTWR 341
DB 446 GGPLCPGSHAVGIFRAVCTRGVAKAVDPFVPSMETTWR 485

```

```

RESULT 15
US-10-191-966-4
; Sequence 4, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; US-10-191-966-4

```

```

Query Match      95.5%; Score 1698; DB 14; Length 1692;
Best Local Similarity 94.1%; Pred. No. 4.6e-165;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 RGRDAIILTCVAHPELIPDITKLLAIFGPMVLOAGITKYPYFVRAQGLIRACMLV 61
DB 146 RGRDAIILTCVAHPELIPDITKLLAIFGPMVLOAGITKYPYFVRAQGLIRACMLV 205
QY 62 KKAAGHYQMAFMKLAALITGYVDHLTPLOMAHAGLRLAVALVEPVFSDMEVKIIT 121
DB 206 KKAAGHYQMAFMKLAALITGYVDHLTPLOMAHAGLRLAVALVEPVFSDMEVKIIT 265
QY 122 GADTAACGDIISGLPVASARRGRIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 181

```

Fri May 7 13:37:05 2004

us-10-650-585-14.rapb

Page 7

Db 266 GADTAACGDILLGPVSARRGKEILLGPADSLEGRGWRLLAPITAVSQOTRGLGCIITS 325  
QY 182 LTGRDKNOVEGEVOVSTATOSPLATCVNGVCMVTFHGAGSKTLAGPKGPITOMYTNVDQ 241  
Db 326 LTGRDKNOVEGEVOVSTATOSPLATCVNGVCMVTFHGAGSKTLAGPKGPITOMYTNVDQ 385  
QY 242 DLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGSLSPRPVSYLKGSSG 301  
Db 386 DLVGMQAPPGARSLTPCTCGSSDLYLVTRHADVI PVRRRGDSRGSLSPRPVSYLKGSSG 445  
QY 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 341  
Db 446 GPLLCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 485

Search completed: May 6, 2004, 09:43:22  
Job time : 35.8831 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 39.7461 Seconds  
(without alignments)  
2075.771 Million cell updates/sec

Title: US-10-650-585-15  
Perfect score: 1532  
Sequence: 1 AAGGIRKCMVLRKAGSHYV.....RGVAKAVDFIPVESMETWR 292

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1532	100.0	292	5	ABG32188 HCV prote
2	1532	100.0	303	5	ABG32183 HCV prote
3	1532	100.0	334	5	ABG32182 HCV prote
4	1532	100.0	341	5	ABG32187 HCV prote
5	1532	100.0	352	5	ABG32186 HCV prote
6	1532	100.0	380	5	ABG32185 HCV prote
7	1532	100.0	393	5	ABG32184 HCV prote
8	1532	100.0	409	5	ABG32181 HCV prote
9	1523	99.4	303	5	ABG32191 HCV prote
10	1522	99.3	303	5	ABG32189 HCV prote
11	1513	98.8	301	5	ABG32190 HCV prote
12	1481	96.7	768	2	AAR40223 Recombina
13	1476	96.3	3010	5	AAE20477 HCV-S1 fu
14	1473	96.1	2201	5	ABG30601 Hepatitis
15	1473	96.1	2201	5	ABG30591 Hepatitis
16	1473	96.1	2201	5	ABG30600 Hepatitis
17	1473	96.1	2201	5	ABG30581 Hepatitis
18	1473	96.1	2201	5	ABG30593 Hepatitis
19	1473	96.1	2201	5	ABG30582 Hepatitis
20	1473	96.1	2201	5	ABG30580 Hepatitis
21	1473	96.1	2201	5	ABG30587 Hepatitis
22	1473	96.1	2201	5	ABG30599 Hepatitis
23	1473	96.1	2201	5	ABG30594 Hepatitis
24	1473	96.1	2201	5	ABG30598 Hepatitis
25	1473	96.1	2201	5	ABG30595 Hepatitis

25	1473	96.1	2307	3	AAY70064 Recombina
27	1473	96.1	3010	2	AAR68822 HCV prote
28	1473	96.1	3010	2	AAR62694 Hepatit H
29	1473	96.1	3010	5	ABG32458 Hepatitis
30	1473	96.1	3010	5	ABG32459 Hepatitis
31	1473	96.1	3010	5	ABG32451 Hepatitis
32	1473	96.1	3010	5	ABG32455 Hepatitis
33	1473	96.1	3010	5	ABG32457 Hepatitis
34	1473	96.1	3010	5	ABG32460 Hepatitis
35	1473	96.1	3010	5	ABG32461 Hepatitis
36	1473	96.1	3010	5	ABG32454 Hepatitis
37	1473	96.1	3011	5	ABG32456 Hepatitis
38	1470	96.0	344	2	AAR62695 HCV parti
39	1470	96.0	2201	5	ABG30586 Hepatitis
40	1470	96.0	2201	5	ABG30589 Hepatitis
41	1470	96.0	2201	5	ABG30583 Hepatitis
42	1470	96.0	2201	5	ABG30588 Hepatitis
43	1470	96.0	2307	3	AAY70065 Recombina
44	1469	95.9	2201	5	ABG30590 Hepatitis
45	1467	95.8	3010	2	AAR68864 Hepatitis

## ALIGNMENTS

RESULT 1  
ID ABG32188 standard; protein; 292 AA.  
XX  
AC ABG32188;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE HCV protease NS2/3 truncation mutant 915-1206.  
XX  
KM HCV, enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection;  
KM Chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KM chaotropic agent; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
OS  
PN WO200248375-A2.  
XX  
PD 20-JUN-2002.  
XX  
PF 13-DEC-2001; 2001MO-CA001796.  
XX  
PR 15-DEC-2000; 2000US-0256031P.  
XX  
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.  
XX  
PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX WPI, 2002-599511/64.  
DR Novel polypeptide for screening inhibitors of non-structural proteases  
XX useful as therapeutic agents against hepatitis C virus, comprises full  
XX length non-structural protease, or its truncation.  
XX  
PS Claim 41; Page 63; 67pp; English.  
XX  
CC The invention relates to an isolated polypeptide consisting of a full-  
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
CC NS2/3 protease. Also included are (1) a composition (c) comprising an  
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
CC its truncation or a mutated sequence, where the protease is in a solution  
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC chaotropic agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products of their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 915-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 292 AA;

Query Match 100.0%; Score 1532; DB 5; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-140;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAGAGHYVQMAFMKLAALGTGVYVYDHLPLQDMAHAGRLDAVAVEPV 60  
 DB 1 AAGIIRACMLVRKAGAGHYVQMAFMKLAALGTGVYVYDHLPLQDMAHAGRLDAVAVEPV 60  
 QY 61 IFSDEVKIITWGDATACGDIISGLPVSARGREIILGPDNFEQGMRLAPITAYSQ 120  
 DB 61 IFSDEVKIITWGDATACGDIISGLPVSARGREIILGPDNFEQGMRLAPITAYSQ 120  
 QY 121 QTRGLGCIITSLTGRKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 180  
 DB 121 QTRGLGCIITSLTGRKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 180  
 QY 181 GPIQMTYNTVDQDLVGQAPPGARSMPTCTGSSDLVYTHAAVYIPRRRGDSRGSILS 240  
 DB 181 GPIQMTYNTVDQDLVGQAPPGARSMPTCTGSSDLVYTHAAVYIPRRRGDSRGSILS 240  
 QY 241 PRPVSYLKGSSGGPILCPSGHNAVIFRAAVCTRGVAKAVDPIPVESMETTR 292  
 DB 241 PRPVSYLKGSSGGPILCPSGHNAVIFRAAVCTRGVAKAVDPIPVESMETTR 292

RESULT 2  
 ABG32198  
 ID ABG32183 standard; protein; 303 AA.

XX ABG32183;

XX 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 904-1206.

XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; viremia;  
 XX hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.

XX Hepatitis C virus.  
 XX Synthetic.

XX MO200248375-A2.

XX 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CA001796.  
 XX  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX  
 XX (BOH) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 XX Thibeault D, Lamarre D, Maurice R, Palocle L, Pause A.  
 XX WPI; 2002-599511/64.  
 XX  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 XX useful as therapeutic agents against hepatitis C virus, comprises full  
 XX length non-structural protease, or its truncation.

XX Claim 39; Page 58-59; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 XX to also as NS2/3 (910-1206)) or its truncation, having as its N-terminal  
 XX residue amino acid 810 to 906, or having a minimal amino acid sequence  
 XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 XX NS2/3 protease. Also included are (1) a composition (C) comprising an  
 XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 XX its truncation or a mutated sequence, where the protease is in a solution  
 XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 XX appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 XX protease, involving isolating the protease in the presence of a  
 XX chaotropic agent, refolding the isolated protease by contacting it with a  
 XX reducing agent, and LDAO in the presence of reduced concentration of the  
 XX chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 XX protease, involving diluting refolded inactive NS2/3 protease in a medium  
 XX containing an activation detergent to induce auto-cleavage of the NS2/3  
 XX protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 XX protease, involving incubating the active NS2/3 protease produced by M2  
 XX for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 XX cleavage products or their fragments, and measuring the presence or  
 XX absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 XX; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 XX active NS2/3 protease, involving carrying out M3 in the presence of, or  
 XX absence of the potential inhibitor, comparing the amount of uncleaved  
 XX NS2/3 protease, cleavage products or their fragments. The protease is  
 XX useful for detailed biochemical characterisation of the enzymes and in  
 XX the development of in vitro assays for screening novel inhibitors of  
 XX NS2/3 protease which are useful as therapeutic agents against HCV  
 XX infection (which causes chronic liver disease, cirrhosis and end-stage  
 XX liver disease. M1 is useful for high level production of protease. The  
 XX present sequence represents the NS2/3 truncation mutant 904-1206  
 XX (numbered relative to the full length NS2/3 protein)

XX Sequence 303 AA;

Query Match 100.0%; Score 1532; DB 5; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-140;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAGAGHYVQMAFMKLAALGTGVYVYDHLPLQDMAHAGRLDAVAVEPV 60  
 DB 1 AAGIIRACMLVRKAGAGHYVQMAFMKLAALGTGVYVYDHLPLQDMAHAGRLDAVAVEPV 71  
 QY 61 IFSDEVKIITWGDATACGDIISGLPVSARGREIILGPDNFEQGMRLAPITAYSQ 120  
 DB 61 IFSDEVKIITWGDATACGDIISGLPVSARGREIILGPDNFEQGMRLAPITAYSQ 131  
 QY 121 QTRGLGCIITSLTGRKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 180  
 DB 121 QTRGLGCIITSLTGRKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 191  
 QY 181 GPIQMTYNTVDQDLVGQAPPGARSMPTCTGSSDLVYTHAAVYIPRRRGDSRGSILS 240  
 DB 181 GPIQMTYNTVDQDLVGQAPPGARSMPTCTGSSDLVYTHAAVYIPRRRGDSRGSILS 251  
 QY 241 PRPVSYLKGSSGGPILCPSGHNAVIFRAAVCTRGVAKAVDPIPVESMETTR 292

Db 252 PRPVSLKSGSGPLLCPSGHAIVGFRAAVCTRGVAKAVDFIPVSMETMR 303

## RESULT 3

ABG32182 ID ABG32182 standard; protein; 334 AA.

AC ABG32182;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.

KM HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;

KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;

KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;

OS chaotropic agent; 4K-6H (904-1206)st-4K; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

FT Key

FT Peptide

FT Protein

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
CC absence of the potential inhibitor, comparing the amount of uncleaved  
CC NS2/3 protease, cleavage products or their fragments. The protease is  
CC useful for detailed biochemical characterisation of the enzymes and in  
CC the development of in vitro assays for screening novel inhibitors of  
CC NS2/3 protease which are useful as therapeutic agents against HCV  
CC infection (which causes chronic liver disease, cirrhosis and end-stage  
CC liver disease. M1 is useful for high level production of protease. The  
CC present sequence represents the NS2/3 truncation mutant 4K-6H (904-  
CC 1206)st-4K comprising a truncated NS2/3 protein with a four Lys/six His N  
CC terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag

XX Sequence 334 AA;

Query Match 100.0%; Score 1532; DB 5; Length 334;

Best Local Similarity 100.0%; Pred. No. 1.4e-140; Mismatches 0; Indels 0; Gaps 0;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIPACMLVRAAGGHHYQVAFMKLALGTYYDHLTPLODMHAGLRLAVALVEPV 60

Db 27 AAGLIPACMLVRAAGGHHYQVAFMKLALGTYYDHLTPLODMHAGLRLAVALVEPV 86

QY 61 IFSDMEVKIITWGADTRACGDIISGLPVASRRGRSTLLGPADNFEQGRLLAPITAYSQ 120

Db 87 IFSDMEVKIITWGADTRACGDIISGLPVASRRGRSTLLGPADNFEQGRLLAPITAYSQ 146

QY 121 QTRGLIGCIITSLTGRDNQVEGEVQVSTATQSPFATCVNGVQWTFHAGSKTLAGRK 180

Db 147 QTRGLIGCIITSLTGRDNQVEGEVQVSTATQSPFATCVNGVQWTFHAGSKTLAGRK 206

QY 181 GPTTQMTNVDDIVQWQAPRGARSMTPTCTGSSDLYVTRADYIPVRRGDSGSLIS 240

Db 207 GPTTQMTNVDDIVQWQAPRGARSMTPTCTGSSDLYVTRADYIPVRRGDSGSLIS 266

QY 241 PRPVSLKSGSGPLLCPSGHAIVGFRAAVCTRGVAKAVDFIPVSMETMR 292

Db 267 PRPVSLKSGSGPLLCPSGHAIVGFRAAVCTRGVAKAVDFIPVSMETMR 318

## RESULT 4

ABG32187 ID ABG32187 standard; protein; 341 AA.

AC ABG32187;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 866-1206.

KM HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;

KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;

KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;

OS chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

PN WO200248375-A2.

PN 20-JUN-2002.

PN 13-DEC-2001; 2001WO-CA001796.

PN 15-DEC-2000; 2000US-0256031P.

PN (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PN Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;

PN WPI; 2002-599511/64.

Novel polypeptide for screening inhibitors of non-structural proteases

PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

PS Claim 41; Page 62-63; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32186; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 866-1206  
 CC (numbered relative to the full length NS2/3 protein)

SQ Sequence 341 AA;

Query Match 100.0%; Score 1532; DB 5; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-140; Indels 0; Gaps 0;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHHYVOMAFMKLAALGTIVYDHLTLPDWMANAGRLDAVAPEV 60  
 DB 50 AAGLIRACMLVRKAGGHHYVOMAFMKLAALGTIVYDHLTLPDWMANAGRLDAVAPEV 109  
 QY 61 IFSMVEKLTITWGDPTAACGDTISGLFVSARRGREILGSPADNPEGQWMLLPITAYSQ 120  
 DB 110 IFSMVEKLTITWGDPTAACGDTISGLFVSARRGREILGSPADNPEGQWMLLPITAYSQ 169  
 QY 121 QTRGLGCIITSLGRBNQVEGEVQVSTATQSFLLATCVNVCWTFHAGAGSKTLAIPK 180  
 DB 170 QTRGLGCIITSLGRBNQVEGEVQVSTATQSFLLATCVNVCWTFHAGAGSKTLAIPK 229  
 QY 181 GPIITMTNVDODIVGQAPPGARSMTPTCGSSDLYLVTRHADYIVRRRGRSGSLLS 240  
 DB 230 GPIITMTNVDODIVGQAPPGARSMTPTCGSSDLYLVTRHADYIVRRRGRSGSLLS 289  
 QY 241 PRPVSYLKGGSGGFLPCPSGAVGIFRAAVCTRGVAAVAFIPVESMETMR 292  
 DB 290 PRPVSYLKGGSGGFLPCPSGAVGIFRAAVCTRGVAAVAFIPVESMETMR 341

RESULT 5

ABG32186 standard; protein; 352 AA.

AC ABG32186;  
 XX 05-NOV-2002 (first entry)  
 DT

XX HCV protease NS2/3 truncation mutant 855-1206.

XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
 KW chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KW hepatocellular carcinoma; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KW chaotropic agent; mutant; mutagen.

OS Hepatitis C virus.  
 OS Synthetic.

PN WO200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001MO-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

PI Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;

DR WPI; 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.

PS Claim 41; Page 61-62; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32186; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 855-1206  
 CC (numbered relative to the full length NS2/3 protein)

SQ Sequence 352 AA;

Query Match 100.0%; Score 1532; DB 5; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-140; Indels 0; Gaps 0;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHHYVOMAFMKLAALGTIVYDHLTLPDWMANAGRLDAVAPEV 60  
 [Sequence alignment lines follow]

```

Db      61 AAGLIRACMLVKAAGHYQVAFMKLAALITGYVYDHLTPLODMAHAGRLDIAVAPEV 120
Qy      61 IFSDEMEKIIITWGAADTAACGDIISGLPVSARRGREIILGPADNFEQGWRLAPITAVSQ 120
Db      121 IFSDEMEKIIITWGAADTAACGDIISGLPVSARRGREIILGPADNFEQGWRLAPITAVSQ 180
Qy      121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSPLATCNGVCMTVPFGAGSKTLAGPK 180
Db      181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSPLATCNGVCMTVPFGAGSKTLAGPK 240
Qy      181 GPITQWNTVNDODLVGMQAPPGARSMPTCTCGSSDILYLVTRHADVIPIVRRRDSRGSLLS 240
Db      241 GPITQWNTVNDODLVGMQAPPGARSMPTCTCGSSDILYLVTRHADVIPIVRRRDSRGSLLS 300
Qy      241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVPIVSEMETTMR 292
Db      301 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVPIVSEMETTMR 352

```

## RESULT 6

ABG32185  
ID ABG32185 standard; protein; 380 AA.

XX ABG32185;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 827-1206.

XX HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.

OS Hepatitis C virus.  
 OS Synthetic.

XX W0200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CA001796.

XX 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases  
 useful as therapeutic agents against hepatitis C virus, comprises full  
 length non-structural protease, or its truncation.

PS Claim 41; Page 60-61; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO),  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium

CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterization of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 827-1206  
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 380 AA;

Query Match 100.0%; Score 1532; DB 5; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-140;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 AAGLIRACMLVKAAGHYQVAFMKLAALITGYVYDHLTPLODMAHAGRLDIAVAPEV 60
Db      89 AAGLIRACMLVKAAGHYQVAFMKLAALITGYVYDHLTPLODMAHAGRLDIAVAPEV 148
Qy      61 IFSDEMEKIIITWGAADTAACGDIISGLPVSARRGREIILGPADNFEQGWRLAPITAVSQ 120
Db      149 IFSDEMEKIIITWGAADTAACGDIISGLPVSARRGREIILGPADNFEQGWRLAPITAVSQ 208
Qy      121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSPLATCNGVCMTVPFGAGSKTLAGPK 180
Db      209 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSPLATCNGVCMTVPFGAGSKTLAGPK 268
Qy      181 GPITQWNTVNDODLVGMQAPPGARSMPTCTCGSSDILYLVTRHADVIPIVRRRDSRGSLLS 240
Db      269 GPITQWNTVNDODLVGMQAPPGARSMPTCTCGSSDILYLVTRHADVIPIVRRRDSRGSLLS 328
Qy      241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVPIVSEMETTMR 292
Db      329 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVPIVSEMETTMR 380

```

## RESULT 7

ABG32184  
ID ABG32184 standard; protein; 393 AA.

XX ABG32184;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 815-1206.

XX HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.

OS Hepatitis C virus.  
 OS Synthetic.

XX W0200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WO-CA001796.

XX 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

PI Thibeault D, Lamarre D, Maurice R, Plote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 XX Claim 41; Page 59-60; 67pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 815-1206  
 CC (numbered relative to the full length NS2/3 protein)  
 XX  
 XX Sequence 393 AA:  
 SQ  
 Query Match 100.0%; Score 1532; DB 5; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1,8e-140;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAGGIRACMLVRRKAGSHVQMAFMKLAALTGTYYVDHLPLDMDMAHGRDLAAVVEPV 60  
 DB 102 AAGGIRACMLVRRKAGSHVQMAFMKLAALTGTYYVDHLPLDMDMAHGRDLAAVVEPV 161  
 QY 61 IFSDMEXKITTWGADTAACDITISGLPVARSRRREILLGPANFEGQWRLLAPITAYSQ 120  
 DB 162 IFSDMEXKITTWGADTAACDITISGLPVARSRRREILLGPANFEGQWRLLAPITAYSQ 221  
 QY 121 QTRGLLCITTSITGRPNKQGEVEVQVSTATQSPFATCNGCVMVFHGAGSKTLAAGR 180  
 DB 222 QTRGLLCITTSITGRPNKQGEVEVQVSTATQSPFATCNGCVMVFHGAGSKTLAAGR 281  
 QY 181 GPITOMTANVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRADVLPVARRGDSRSLLS 240  
 DB 282 GPITOMTANVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRADVLPVARRGDSRSLLS 341  
 QY 241 PRVSVSYLKSGSGGPLICPSGHAVGIFRAAVCTRGVAKADVLPVESMETTMR 292  
 DB 342 PRVSVSYLKSGSGGPLICPSGHAVGIFRAAVCTRGVAKADVLPVESMETTMR 393

ID ABG32181 standard; protein; 409 AA.  
 XX  
 XX ABG32181;  
 AC  
 XX  
 XX 05-NOV-2002 (first entry)  
 DT  
 XX  
 XX HCV protease NS2/3 (810-1206).  
 DE  
 XX  
 XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;  
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 XX hepatotropic; antinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutcin.  
 XX  
 XX Hepatitis C virus.  
 OS  
 XX Synthetic.  
 FT  
 XX Key location/Qualifiers  
 FT Peptide 398..409  
 FT /note="Streptavidin tag"  
 XX  
 XX WO200248375-A2.  
 XX  
 XX 20-JUN-2002.  
 XX  
 XX 13-DEC-2001; 2001WO-CR001796.  
 XX  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX  
 XX (BOER) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 XX Thibeault D, Lamarre D, Maurice R, Plote L, Pause A;  
 XX WPI; 2002-599511/64.  
 XX N-PSDB; ABK90406.  
 DR  
 XX  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 XX Claim 42; Fig 1b; 67pp; English.  
 PS  
 XX  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 (810-1206) protein, which has a C-

CC terminal streptavidin tag  
 XX Sequence 409 AA;  
 SQ  
 Query Match 100.0%; Score 1532; DB 5; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-140;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIACMLVRRKAGGHVQMAFMKLAALGTYYVDHLTPLODMHAGLRDLAVAVEPV 60  
 DB 106 AAGLIACMLVRRKAGGHVQMAFMKLAALGTYYVDHLTPLODMHAGLRDLAVAVEPV 165  
 QY 61 IFSDEVKIITWGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSQ 120  
 DB 166 IFSDEVKIITWGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSQ 225  
 QY 121 QTRGLIGCIITSLTGRDKNQEVEGVVSTATOSPLATCNGVCWTFHAGSKTLAGPK 180  
 DB 226 QTRGLIGCIITSLTGRDKNQEVEGVVSTATOSPLATCNGVCWTFHAGSKTLAGPK 285  
 QY 181 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVRRRGDSRGLLS 240  
 DB 286 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVRRRGDSRGLLS 345  
 QY 241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292  
 DB 346 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 397

RESULT 9  
 ABG32191  
 ID ABG32191 standard; protein; 303 AA.  
 AC ABG32191;  
 DT 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation 904-1206/Cys93Aa.  
 DE HCV protease NS2/3 truncation 904-1206/Cys93Aa.  
 XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 KM chaotropic agent; mutant; mutein.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH Misc-difference 90 /note="Wild-type Cys substituted by Ala"  
 FT  
 XX WO200248375-A2.  
 XX 20-JUN-2002.  
 XX 13-DEC-2001; 2001WO-CA001796.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibault D, Lamare D, Maurice R, Pilote L, Pause A;  
 XX WPI, 2002-599511/64.  
 XX Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX Disclosure; Page 65-66; 67pp; English.  
 XX The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred

CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of the protease. The  
 CC present sequence represents the NS2/3 truncation 904-1206 mutant  
 CC Cys93Aa (numbered relative to the full length NS2/3 protein) a mutant  
 CC devoid of autocatalytic activity  
 CC  
 XX  
 SQ Sequence 303 AA;  
 Query Match 99.4%; Score 1523; DB 5; Length 303;  
 Best Local Similarity 99.7%; Pred. No. 9.3e-140;  
 Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGLIACMLVRRKAGGHVQMAFMKLAALGTYYVDHLTPLODMHAGLRDLAVAVEPV 60  
 DB 12 AAGLIACMLVRRKAGGHVQMAFMKLAALGTYYVDHLTPLODMHAGLRDLAVAVEPV 71  
 QY 61 IFSDEVKIITWGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSQ 120  
 DB 72 IFSDEVKIITWGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSQ 131  
 QY 121 QTRGLIGCIITSLTGRDKNQEVEGVVSTATOSPLATCNGVCWTFHAGSKTLAGPK 180  
 DB 132 QTRGLIGCIITSLTGRDKNQEVEGVVSTATOSPLATCNGVCWTFHAGSKTLAGPK 191  
 QY 181 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVRRRGDSRGLLS 240  
 DB 192 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPVRRRGDSRGLLS 251  
 QY 241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292  
 DB 252 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 303

RESULT 10  
 ABG32189  
 ID ABG32189 standard; protein; 303 AA.  
 AC ABG32189;  
 DT 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation 904-1206/His95Aa.  
 DE HCV protease NS2/3 truncation 904-1206/His95Aa.  
 XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;  
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;

KM chaotropic agent; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 49 /note="Wild-type His substituted by Ala"  
 FT  
 XX  
 EN WO200248375-A2.  
 PD 20-JUN-2002.  
 XX  
 PD 13-DEC-2001; 2001WO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 PS Example 7; Fig 8; 67pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation 904-1206 mutant  
 CC H1595A1a (numbered relative to the full length NS2/3 protein) a mutant  
 CC devoid of autocatalytic activity  
 XX  
 SQ Sequence 303 AA.  
 Query March 99.3%; Score 1522; DB 5; Length 303;  
 Best Local Similarity 99.7%; Pred. No. 1.2e-139;  
 Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAGTAAACMKYKAGAGHYQMAFMALATCTGYVDPHTPLPDVMAHAGRLAVAVEV 60  
 DB 12 AAGTAAACMKYKAGAGHYQMAFMALATCTGYVDPHTPLPDVMAHAGRLAVAVEV 71

QY 61 IFSDMEVKIITWGAADTAACGDIISGLPVASARGREILIGPADNFEQGMRLAPITANSQ 120  
 DB 72 IFSDMEVKIITWGAADTAACGDIISGLPVASARGREILIGPADNFEQGMRLAPITANSQ 131  
 QY 121 QTRGLGCIITSLTRDRKNQVGEVQVSTATQSFATCNAVGCWTVHAGSKTLGAPK 180  
 DB 132 QTRGLGCIITSLTRDRKNQVGEVQVSTATQSFATCNAVGCWTVHAGSKTLGAPK 191  
 QY 181 GPITQMTNVDPDLVGMQAPPGARSMPTCTGSSDLYVTEHADVIYRRRDSRGSLIS 240  
 DB 192 GPITQMTNVDPDLVGMQAPPGARSMPTCTGSSDLYVTEHADVIYRRRDSRGSLIS 251  
 QY 241 PRPVSYLKGSGGGLPCPSGHAIVGFRAAVCTRGVAKAVDPFVPSMETTKR 292  
 DB 252 PRPVSYLKGSGGGLPCPSGHAIVGFRAAVCTRGVAKAVDPFVPSMETTKR 303  
 RESULT 11  
 ABG32190  
 ID ABG32190 standard; protein; 301 AA.  
 XX  
 AC ABG32190;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE HCV protease NS2/3 truncation 904-1206/deltaLeu1026-A1a1027.  
 XX  
 XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;  
 KM chronic liver disease; cirrhosis; end-stage liver disease; virocidic;  
 KM hepatocarcinoma; antiinflammatory; lauryldiethylamine oxide; LDAO;  
 XX chaotropic agent; mutant; mutein.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 122.123  
 FT  
 XX  
 PN WO200248375-A2.  
 PD 20-JUN-2002.  
 XX  
 PF 13-DEC-2001; 2001WO-CA001796.  
 XX  
 PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 DR WPI; 2002-599511/64.  
 XX  
 PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 XX  
 PS Example 7; Page 64-65; 67pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal  
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide  
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a



reducing agent, and DDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 CC protease, involving incubating the active NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease. M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation 904-1206 mutant  
 CC delta1026-1027 (numbered relative to the full length NS2/3 protein)  
 CC a mutant devoid of autocatalytic activity

Sequence 301 AA:

Query Match 98.8%; Score 1513; DB 5; Length 301;  
 Best Local Similarity 99.3%; Pred. No. 8.7e-139;  
 Matches 290; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

1 AAGLRACMLVRKAGHYVQAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 60  
 12 AAGLRACMLVRKAGHYVQAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 71  
 QY 1 FSDMEVKITITGADTAACGDIISGLPVSARSGEILLGPADNFEQGMRLAPITAYSQ 120  
 Db 72 IFSMEVKITITGADTAACGDIISGLPVSARSGEILLGPADNFEQGMRL--PIITAYSQ 129  
 QY 121 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGASKTLAEPK 180  
 Db 130 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGASKTLAEPK 189  
 QY 181 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLLS 240  
 Db 190 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLLS 249  
 QY 241 PRPVSYLKSGSGGPIPCPSGHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 292  
 Db 250 PRPVSYLKSGSGGPIPCPSGHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 301

RESULT 12

AAR40223 standard; protein; 768 AA.

AAR40223;

21-FEB-1994 (first entry)

Recombinant hepatitis C virus genomic protein.

Hepatitis C virus; screening; antiviral drugs.

Hepatitis C virus.

Hepatitis C virus.

Key Location/Qualifiers

Misc-difference 10 /note= "UUA encodes Ile."

Misc-difference 51.82 /note= "Nucleotide sequence encodes another Gly"

Duplication 528.768 /note= "Duplication of 241 amino acids at start of protein sequence."

Misc-difference 537 /note= "UUA encodes Ile."

JP05192160-A.  
 03-AUG-1993.  
 20-JAN-1992; 92JP-00028833.  
 20-JAN-1992; 92JP-00028833.  
 (BANYU ) BANYU PHARM CO LTD.  
 WPI: 1993-277474/35.  
 N-PSDB; AA048215.

Hepatitis C virus genomic RNA, cDNA and polypeptide - used for screening hepatitis C virus-specific antiviral drugs.

Claim 9; Page 4-6; 14pp; Japanese.

The protein is useful in the screening of HCV-specific antiviral drugs.

HCV cDNA was cloned from plasma. Plasmids pSR3241 and pSR2541 were prepared using the cDNA and plasmid pSR3241 was used to transform a COS-1 cell

Sequence 768 AA:

Query Match 96.7%; Score 1481; DB 2; Length 768;  
 Best Local Similarity 95.2%; Pred. No. 4.3e-135;  
 Matches 278; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

1 AAGLRACMLVRKAGHYVQAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 60  
 176 AAGLRACMLVRKAGHYVQAFMKLAALTGYVDHLTPLODMAHAGLRDLAVAVEPV 235  
 QY 61 FSDMEVKITITGADTAACGDIISGLPVSARSGEILLGPADNFEQGMRLAPITAYSQ 120  
 Db 236 VFSAMETKITITGADTAACGDIISGLPVSARSGEILLGPADNFEQGMRLAPITAYSQ 295  
 QY 121 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGASKTLAEPK 180  
 Db 296 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGASKTLAEPK 355  
 QY 181 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLLS 240  
 Db 356 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGSLLS 415  
 QY 241 PRPVSYLKSGSGGPIPCPSGHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 292  
 Db 416 PRPVSYLKSGSGGPIPCPSGHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 467

RESULT 13

AAR20477 standard; protein; 3010 AA.

AAR20477;

01-JUL-2002 (first entry)

HCV-S1 full-length polypeptide.

Nucleic acid construct; expression cassette; non-coding region; NCR;

untranslated region; UTR; anti-viral drug; drug resistance; HCV-S1;

Hepatitis C virus.

Hepatitis C virus.

WO200208447-A2.

31-JAN-2002.

20-JUL-2001; 2001WO-11000669.

24-JUL-2000; 2000US-0220248P.

(MOLE-) INST MOLECULAR & CELL BIOLOGY.  
(EHRICH) EHRICH G.

Tan YH, Lim SG, Hong WJ;

WPI: 2002-280605/32.

N-PSDB; MAD33038.

Novel nucleic acid construct useful for detecting the presence of RNA virus, comprises an expression cassette and a promoter operably linked to expression cassette for minus strand RNA transcription of the cassette.

Example 1; Page 70-81; 81pp; English.

The invention relates to nucleic acid construct which comprises an expression cassette including a first polynucleotide region including a 5' non-coding region (NCR) sequence of an RNA virus and at least an N-terminal portion of a coding sequence of RNA virus, a second polynucleotide region including a 3' untranslated region (UTR) sequence of the RNA virus and at least a C-terminal portion of a coding sequence of the virus and a third polynucleotide region encoding a reporter molecule, flanked by first and second polynucleotide regions; and a promoter sequence being operatively linked to expression cassette in a manner so as to enable a transcription of a minus strand RNA molecule from the expression cassette. Nucleic acid construct of the invention is useful for detecting the presence of an RNA virus in a cell. It is also useful for screening anti-viral drugs and determining drug resistance of an RNA virus. The present sequence is Hepatitis C virus (HCV) isolate HCV-S1 full-length polypeptide.

Sequence 3010 AA;

Query Match 96.3%; Score 1476; DB 5; Length 3010;

Best Local Similarity 95.2%; Pred. No. 8.9e-134; Indels 0; Gaps 0;

Matches 278; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

1 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60

915 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 974

61 IFSMEVKIITWADTAACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 120

975 VFSMEVKIITWADTAACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 1034

121 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180

1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 1094

181 GPITOMTNNVDQIVGQAPPGARSMTPTCTGSSDLYLVRHADVIIVRRRGDSRGLLS 240

1095 GPITOMTNNVDQIVGQAPPGARSMTPTCTGSSDLYLVRHADVIIVRRRGDSRGLLS 1154

241 PRPVSYLKSGSGGFLCPSGHAGVIFRAAVCTRGVAKADVIFVESMETMR 292

1155 PRPVSYLKSGSGGFLCPSGHAGVIFRAAVCTRGVAKADVIFVESMETMR 1206

RESULT 14

ABG30601 standard; protein; 2201 AA.

ABG30601;

21-OCT-2002 (first entry)

Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.

Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.

Hepatitis C virus.

Synthetic.

Key Location/Qualifiers

Misc-difference 882

FT /label= Arg, Lys

Misc-difference 2183 /note= "Wild type Met substituted by Thr"

WO200252015-A2.

04-JUL-2002.

20-DEC-2001; 2001WO-CA001843.

22-DEC-2000; 2000US-0257857P.

WPI: 2002-575382/61.

Kukolj G, Pause A;

New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.

Claim 3; Page; 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polypeptide region coding for a HCV polypeptide; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon Apk12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention

Sequence 2201 AA;

Query Match 96.1%; Score 1473; DB 5; Length 2201;

Best Local Similarity 94.9%; Pred. No. 1.1e-133; Indels 0; Gaps 0;

Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

1 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60

106 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 165

61 IFSMEVKIITWADTAACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 120

166 VFSMEVKIITWADTAACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 225

121 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180

226 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 285

181 GPITOMTNNVDQIVGQAPPGARSMTPTCTGSSDLYLVRHADVIIVRRRGDSRGLLS 240

286 GPITOMTNNVDQIVGQAPPGARSMTPTCTGSSDLYLVRHADVIIVRRRGDSRGLLS 345

241 PRPVSYLKSGSGGFLCPSGHAGVIFRAAVCTRGVAKADVIFVESMETMR 292

346 PRPVSYLKSGSGGFLCPSGHAGVIFRAAVCTRGVAKADVIFVESMETMR 397



